




Machine learning that predicts well may not learn the correct physical descriptions of glassy systemsArabind Swain *Department of Physics, Emory University, Atlanta, Georgia 30322, USA*Sean Alexander Ridout *Department of Physics, Emory University, Atlanta, Georgia 30322, USA
and Initiative in Theory and Modeling of Living Systems, Emory University, Atlanta, Georgia 30322, USA*Ilya Nemenman *Department of Physics, Emory University, Atlanta, Georgia 30322, USA;
Department of Biology, Emory University, Atlanta, Georgia 30322, USA;
and Initiative in Theory and Modeling of Living Systems, Emory University, Atlanta, Georgia 30322, USA*

(Received 4 March 2024; accepted 9 July 2024; published 23 July 2024)

The complexity of glasses makes it challenging to explain their dynamics. Machine learning (ML) has emerged as a promising pathway for understanding glassy dynamics by linking their structural features to rearrangement dynamics. Support vector machine (SVM) was one of the first methods used to detect such correlations. Specifically, a certain output of SVMs trained to predict dynamics from structure, the distance from the separating hyperplane, was interpreted as being linearly related to the activation energy for the rearrangement. By numerical analysis of toy models, we explore under which conditions it is possible to infer the energy barrier to rearrangements from the distance to the separating hyperplane. We observe that such successful inference is possible only under very restricted conditions. Typical tests, such as the apparent Arrhenius dependence of the probability of rearrangement on the inferred energy and the temperature, or high cross-validation accuracy do not guarantee success. Since even in such relatively simple toy models, prediction success of ML models does not necessarily translate into success of learning the underlying physics, we suggest that more careful investigations are needed when such claims are made. For this, we propose practical approaches for measuring the quality of the energy inference and for modifying the inferred model to improve the inference, which should be usable in the context of realistic datasets.

DOI: [10.1103/PhysRevResearch.6.033091](https://doi.org/10.1103/PhysRevResearch.6.033091)**I. INTRODUCTION**

In recent years, there have been a number of attempts to use machine learning (ML) techniques to better understand physical phenomena [1]. One of the areas that has shown considerable promise is the use of classification algorithms to differentiate between different states of a physical system [2–19]. In some of these cases, ML techniques manage to go beyond classification, extracting physically interpretable low-dimensional descriptions, such as order parameters [2,12,17], topological invariants [20], or the energy barriers that determine the rate of rearrangements in a glassy liquid [8,9,18]. In other words, sometimes ML methods build accurate *physical* models of the studied system, even when the relevant variables describing the physics are not explicitly in the dataset. Traditionally, finding such low-dimensional, relevant descriptions requires specialized knowledge, e.g., of

conservation laws. Such successes without this specialized knowledge show the potential of ML techniques to discover new physics with minimal guidance by scientists. However, very little is known about when an ML method, trained to predict a certain aspect of the behavior of a physical system, constructs an accurate physical model, rather than a purely statistical one.

We will answer this question in a simplified, tractable model of the important physical problem of predicting rearrangements of glassy liquids using structural data [8,9,18]. Glassy liquids have heterogeneous rearrangement dynamics: in some regions particles rearrange quickly, while others are slow. The degree of heterogeneity, as well as length scales characterizing the range of dynamical correlations, grows as the temperature is lowered [21–24]. Despite this, the structural order in a glass is hard to detect, making the origin of these correlations difficult to understand [25,26]. In recent years, there has been considerable progress in linking the dynamics of glassy liquids to their structure using ML. Support vector machines (SVMs) [8,9,13,14,18,19,27], neural networks [15,28–32], and linear regression [16,32,33] have been trained on large datasets generated through simulations. Local structural features were used to predict whether a

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

particle rearranges in a specific time period Δt . All of these methods were shown to predict rearrangements with high accuracy. The classifiers could also predict rearrangements when applied to data from previously unseen temperatures. Thus, the classifiers learn local structural predictors of dynamics that generalize across temperatures. In the linear SVM case, the distance to the separating hyperplane, named softness S [8], has a simple interpretation as a local energy barrier to rearrangement $\Delta E(S)$. This is because the probability for a particle to rearrange in some unit time Δt given S was numerically found to obey the Arrhenius law,

$$P(R|S) \propto \exp[\Sigma(S) - \Delta E(S)/T], \quad (1)$$

which is precisely the probability of rearrangement for a process that requires crossing a single energy barrier $\Delta E(S)$. In particular, $\Sigma(S)$ and $\Delta E(S)$ were found to be linear in S . Therefore, this simple linear classifier seems to have learned a physical description of the system, without being instructed to infer it.

Recent work has begun to use this learned dynamical description as the basis for simplified dynamical models of supercooled liquids and amorphous solids, using the inferred $\Delta E(S)$ and $\Sigma(S)$ as parameters in these models [34–37]. However, there has been no explicit study showing if the success in making predictions signifies that the inferred physical description agrees with the true one. Understanding when the two match is the goal of this work. Specifically, assuming that there exists an underlying structural variable S such that Eq. (1) holds in a glassy liquid, we will explore when an SVM can learn the correct variable S . We focus on SVMs [38] (and, more specifically, linear SVMs) in our study because SVMs are interpretable, their performance compares well to other methods for this system, and the interpretation of statistical properties of the classifier (softness) as a physical quantity (linearly proportional to the Arrhenius energy barrier) was made for SVMs, and not other ML methods.

We devise a toy model where a true energy barrier $\Delta E(\vec{x})$ describes the probability for a given configuration \vec{x} to rearrange. We show numerically how the choice of structural variables given to the SVM affects the prediction accuracy and the ability of the trained model to predict the true energy barrier. We show that, if the SVM is given as the input only those features that contribute linearly to $\Delta E(\vec{x})$, then the inferred softness (distance to the separating hyperplane) indeed predicts the true $\Delta E(\vec{x})$. This is true even when the SVM is only trained to predict rearrangements, rather than $\Delta E(\vec{x})$ explicitly. However, we also show that, with a finite amount of training data, the energy barrier estimated through the softness inferred by the SVM can be strongly biased. Surprisingly, this is true even if the quality of prediction, measured by common statistical tests, such as cross-validation, is high. Thus, for our simple model, SVM does not necessarily learn the correct energy barriers, even when it seems that it does or should. Since, in real systems, structural variables determining the energy barrier are typically unknown, one usually provides an ML algorithm with a large set of features, with only some of the features that can act as predictors of the rearrangement probability [8]. One then hopes that the machine distinguishes the features that directly contribute to the barrier height from those that are correlated with them, and from those that are

irrelevant for the prediction. In this scenario, we show that the SVM becomes confused, so that its softness cannot be interpreted as the barrier in the presence of additional features correlated with components of the true energy function. Although the models we study are simple toy models, the fact that SVMs can fail to infer the true energy barriers even in a simple model suggests that their applications in real physical systems should be more carefully tested. Finally, we demonstrate methods to diagnose these problems and to fix them by systematic pruning of the structural features used to predict rearrangements.

II. MODEL AND SIMULATIONS

We study a toy model, which still contains many of the features relevant for our analysis. In the previous work, Ref. [8], an SVM was used to identify a linear combination $S_i = S(\vec{x}_i) = \sum_{j=1}^n \alpha^j x_i^j$ of structural features \vec{x}_i , associated with a specific particle i , such that the probability of rearrangement for the particle is as in Eq. (1). Specifically, in order to reproduce Eq. (1), we require a model where (1) each particle i is described by n structural variables $\vec{x}_i = \{x_i^1, x_i^2, \dots, x_i^n\}$, which vary among the particles; (2) each particle has a rearrangement energy barrier $\Delta E(\vec{x}_i)$, and (3) the probability to rearrange depends on T and $\Delta E(\vec{x}_i)$ with a law that tends to the Arrhenius law for low temperatures. Additionally, we investigated data from simulations from Ref. [36] and found that the distributions of the predictors and the inferred energy were largely Gaussian (see Appendix A). The simplest model with these properties is one where all n dimensions of \vec{x}_i are drawn independently at random, and the true energy barrier is a linear function of the n -dimensional \vec{x}_i . Thus, for each particle $i = 1, \dots, N$, we generate an n -dimensional coordinate vector $\vec{x}_i = \{x_i^1, x_i^2, \dots, x_i^n\}$ as

$$x_i^j \sim \mathcal{N}(0, (\sigma^j)^2) \quad \forall \quad j = 1, \dots, n \quad \text{and} \quad i = 1, \dots, N. \quad (2)$$

We then assume that the energy barrier to rearrangement is a linear combination of these coordinates

$$\Delta E(\vec{x}_i) = \sum_{j=1}^n \alpha^j x_i^j. \quad (3)$$

This results in a Gaussian distribution of ΔE , consistent with the Gaussian distribution of S in supercooled liquids [8].

Finally, for each configuration, we determine whether or not it rearranges by sampling a binary random variable $R_i = \pm 1$ (where ± 1 stands for the presence or absence of a rearrangement) from

$$P(R_i = 1 | \vec{x}_i) = \frac{e^{-\beta \Delta E(\vec{x}_i)}}{1 + e^{-\beta \Delta E(\vec{x}_i)}}, \quad (4)$$

which reduces to the Arrhenius form at low T while remaining below 1 at high T .

We then train a linear SVM [39] to predict R_i from \vec{x}_i , for all $i = 1, \dots, N$. As is the common practice, for the training, we standardize all x s to have zero mean and unit variance. Thus, drawing x^j from $\mathcal{N}(0, (\sigma^j)^2)$ is equivalent to drawing them from $\mathcal{N}(0, 1)$ and absorbing the standard deviation into the definition of α , which is what we do. Further, the results

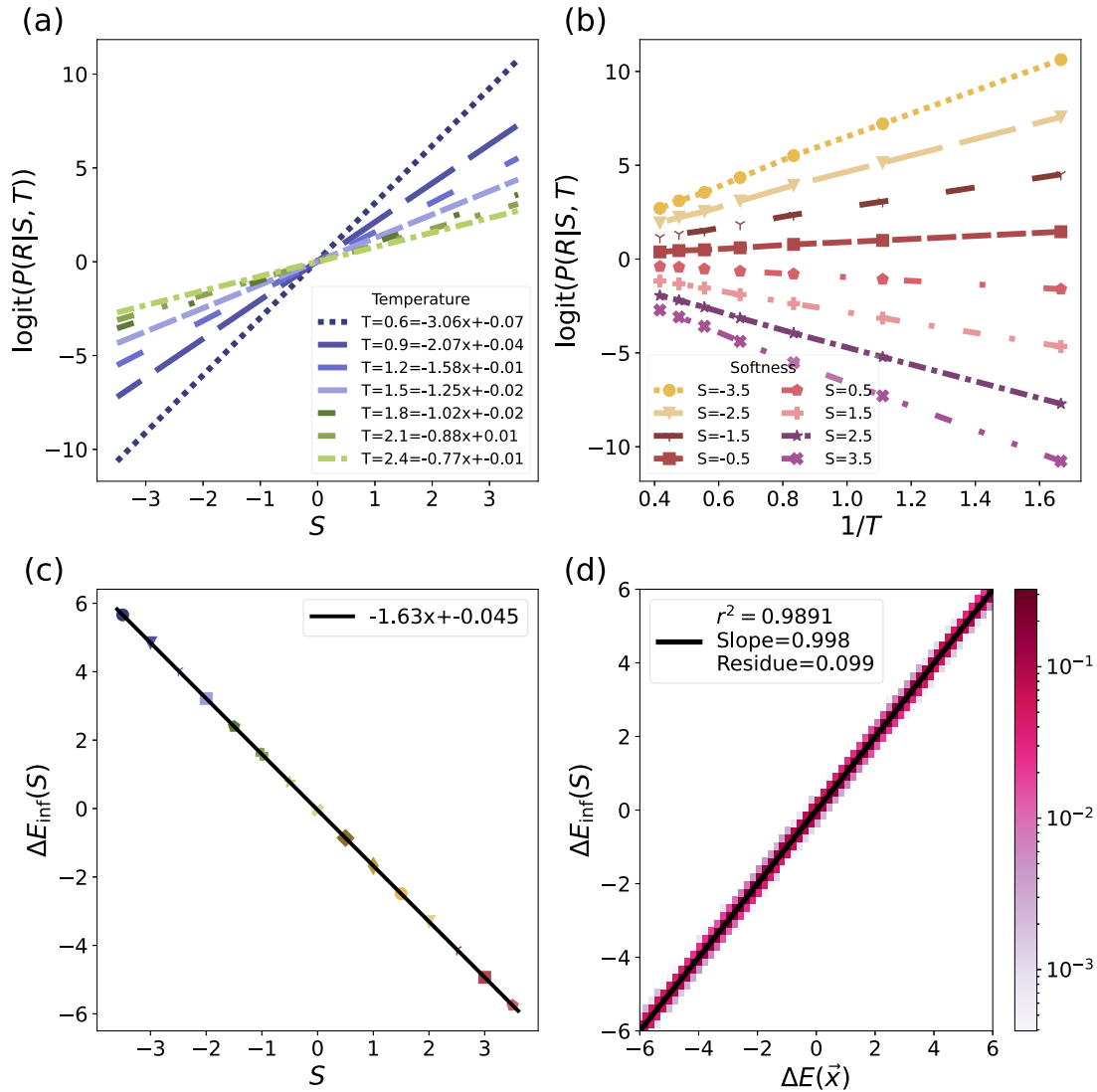


FIG. 1. Relationship between softness S and $\Delta E(\vec{x})$ for symmetric distribution of training energies for a large training set size, $N = 10^6$. (a) $\text{logit} P(R|S)$ derived from fitting the logistic curve to the probability of rearrangement as a function of S for different temperatures T . (b) $\text{logit} P(R|S, T)$ vs $1/T$ for 15 different values of softness. (c) The inferred $\Delta E_{\text{inf}}(S)$, calculated from $\text{logit} P(R|S, T)$, as a function of S . (d) Two-dimensional (2D) joint density plot and the linear fit of the true energy barrier $\Delta E(\vec{x})$ vs the inferred energy barrier $\Delta E_{\text{inf}}(S)$ (we plot the joint density instead of the scatter for clarity of the visualization).

shown below are all evaluated at $\alpha^j = 1.2$. We verified separately that this choice does not change qualitative results from Secs. III and IV (not shown, but also see Appendix B for some discussion).

After training the SVM, we define the *softness* S_i for state \vec{x}_i as the signed distance to the separating hyperplane, as in previous work [8]. We then want to estimate the probability of rearrangement $P(R|S)$, to see if the softness defines it well. In Ref. [8], this probability was estimated as the frequency of rearrangements in a certain small bin of S . Instead, to remove artifacts caused by the finite bin width, we estimate $P(R|S)$ using a logistic regression model.

In a glass, energy barriers should be strictly positive, and the probability for a typical particle to rearrange is tiny. To remove biases in the inference, one typically balances the dataset used for training to have similar numbers of particles

that do and do not rearrange [8]. In our model, Eq. (3), we achieve this balance by explicitly centering ΔE at zero. We checked numerically that this choice does not qualitatively affect the ability of the SVM to correctly predict the energy (Appendix B).

A large number of structural features are used to train an SVM to predict glassy dynamics [8]. These features, however, are correlated. To observe the effect of these correlations on the ability of the SVM to predict the correct energy, for simulations in Sec. V, we give as input to the SVM a $2n$ -dimensional coordinate vector (\vec{x}_i, \vec{z}_i) , where $z_i^j = (x_i^j)^2 \sum_{j_1} x_i^{j_1}$, and all x 's remain uncorrelated, as before. There is nothing particular about this choice of additional variables z^j correlated with x^j , besides that we wanted to preserve the same symmetry under parity (even-order contributions would average out for symmetric x 's). Further, we wanted these

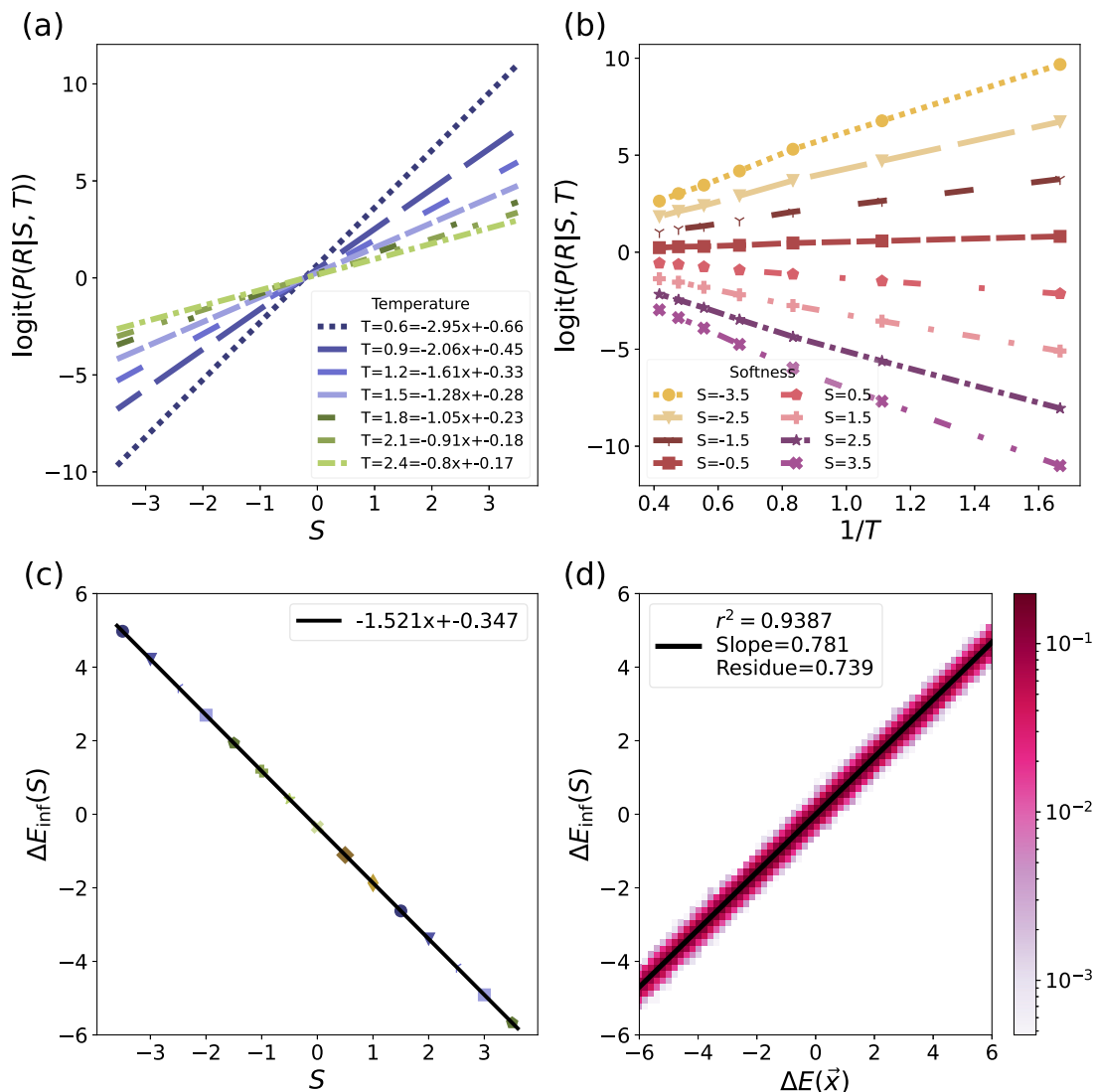


FIG. 2. Relationship between softness and $\Delta E(\vec{x})$ for symmetric distribution of training energies for a small training set size, $N = 10^3$. (a)–(c) Same as in Fig. 1. (d) The true energy barrier $\Delta E(\vec{x})$ vs the inferred energy barrier from SVM $\Delta E_{\text{inf}}(S)$. Note that, to the extent that the slope in (d) is not 1, the correct energy is not learned.

spurious extra dimensions to be nonlinearly correlated with x 's, modeling nonlinear correlations between values of different radial and angular density functions in Ref. [8]. We believe that our conclusions on the ability of the SVM to predict the correct energy will be qualitatively the same for other choices of spurious correlated variables obeying these conditions, and we have checked a few other cases (Appendix D). We then train the SVM to predict rearrangements from this expanded set of coordinates and evaluate the effect of the correlated input variables on the quality of the model the SVM builds.

III. LINEAR SVM CAN LEARN THE TRUE ENERGY BARRIER IN THE INFINITE DATA LIMIT

First, we test whether or not the softness S , inferred by the SVM from a very large sample, is a good approximation for $\Delta E(\vec{x})$ from Eq. (3). We use $N = 10^6$ training samples with 5×10^5 examples each of rearranging and nonrearranging configurations to train the SVM. The distribution

of energies in the training sample is symmetric. We have 14 independently sampled input dimensions, with $\alpha^j = 1.2$ for $j = 1, \dots, 10$ and $\alpha^j = 0$ for $j = 11, \dots, 14$. Thus, ten dimensions determine the energy, while the other four dimensions can be seen as Gaussian noise uncorrelated with any of the relevant input dimensions.

In Fig. 1(a), we show the relationship between the probability for particles to rearrange, $P(R|S)$, and S by plotting $\text{logit } P(R|S) \equiv \log[P(R|S)/(1 - P(R|S))]$ vs S . $P(R|S)$ is calculated by fitting a logistic regression that predicts whether a particle is rearranging from its S . This plot is analogous to the $\log P(R|S)$ vs S plots in earlier studies [8] since in our model $\text{logit } P(R|\Delta E)$ is linear in ΔE . The plot shows a similar linear relationship between $\text{logit } P(R|S)$ and S . When $\text{logit } P(R|S)$ is plotted as a function of $1/T$ for several values of softness [Fig. 1(b)], we also see a linear relationship between $\text{logit } P(R|S)$ and $1/T$ as observed in earlier studies [8]. As in the previous work [8], the slope of $\text{logit } P(R|S)$ vs $1/T$ for each softness S is used to infer the corresponding energy

barrier $\Delta E_{\text{inf}}(S)$ in Fig. 1(c). This $\Delta E_{\text{inf}}(S)$ is analogous to the barrier energy $\Delta E(S)$ in the Arrhenius rate equation, Eq. (1). As one can see, the inferred barrier energy, $\Delta E_{\text{inf}}(S)$, has a linear relationship with softness, S . Thus, our model, in this limit, reproduces the observations of previous work [8]: the probability of rearrangement is exponential in the distance S to the separating hyperplane, a.k.a. softness, and this distance has an interpretation as an inferred energy barrier $\Delta E_{\text{inf}}(S)$.

Unlike in past work, in our model, the true energy barriers are *known*. Thus, we then can compare the inferred energy barrier $\Delta E_{\text{inf}}(S)$ to the true energy barrier $\Delta E(\bar{x})$ for each configuration \bar{x}_i in the test set. We plot the inferred energy vs the true energy, as well as a linear regression line between the two in Fig. 1(d). Since the slope of the fit is ~ 1.0 and the scatter around the linear fit is small, we conclude that the SVM indeed learns the real energy barrier $\Delta E(\bar{x})$ with a high degree of accuracy. We also find that the SVM captures the real energy when trained on unsymmetrical data where all energy barriers are positive (see Appendix B).

IV. LARGE TRAINING SETS ARE REQUIRED FOR SVM TO LEARN TRUE ENERGY BARRIERS

For real-world problems, we do not have access to an infinite (extremely large) amount of data. Thus, it is natural to ask whether inferred energies are still accurate for smaller training sets. For this, we repeated the analysis of Sec. III with varied training set size $N = 10^3, \dots, 10^6$.

As shown in Fig. 2(a)–2(c), when $N = 10^3$, the inference procedure still seems to work. That is, $\log P(R|S, T)$ is still a linear function of S , and it still appears to be linear in $1/T$. This allows us again to infer the energy barrier $\Delta E_{\text{inf}}(S)$, which is linear in S . However, regressing $\Delta E_{\text{inf}}(S)$ against the true $\Delta E(\bar{x})$ shows that the inferred energy is *biased*, consistently underestimating the magnitude of the true energy by nearly 15%. Since the variance of the distribution of true energy $P(\Delta E)$ is a sum of the variance explained by S and the variance unexplained by S , the error must always have this sign: if the energy is inferred incorrectly, the variance of the distribution of inferred energies will be less than the variance of the distribution of true energies. This point is discussed further in Sec. V.

Figure 3(a) shows how this underestimation depends on N . Further, Fig. 3(b) shows the N dependence of the classification (rearranged or not) prediction accuracy of our fitted model on a test set, different from the training one. To verify that fitting and prediction errors do not come from suboptimal choices during training, in this figure, we also change the value of the SVM training hyperparameter C , which controls when the SVM treats data points that are labeled differently from their neighbors as outliers vs true data that should be fitted [38,39]. For small N , regardless of C , the true energy is underestimated. For large N , the quality of the fits improves, and the prediction accuracy as well as the error in slope become largely insensitive to C .

In practice, the true energy is rarely known. Thus, detection of the bias shown in Figs. 2(d) and 3 is nontrivial in experimental applications. Indeed, simple checks, such as verifying the linearity of plots in Figs. 2(a)–2(c), do not reveal this error. Further, the underestimation of the barrier magnitude is also

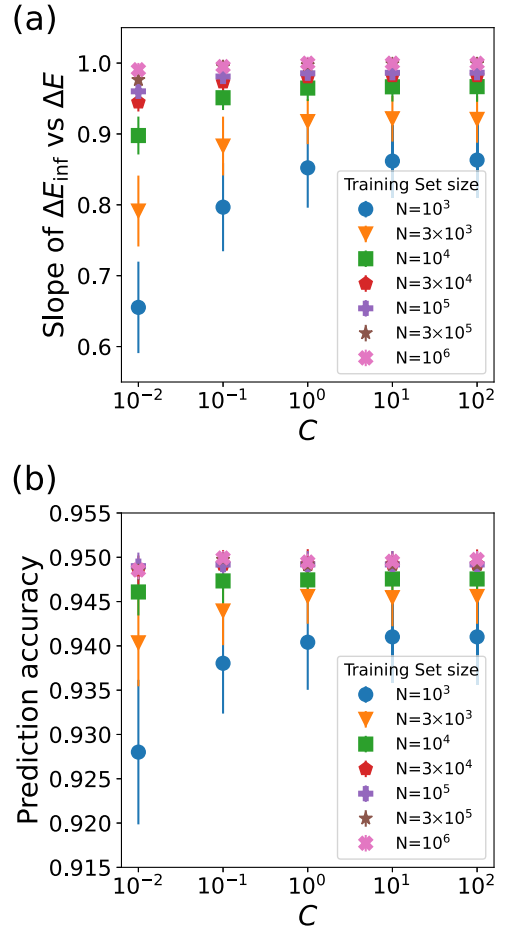


FIG. 3. Slope of inferred energy (a) $\Delta E_{\text{inf}}(S)$ vs real energy $\Delta E(\bar{x})$ and the prediction accuracy (b) for different sizes of training data as a function of the SVM cost parameter C . The training and test data were generated at $T = 0.4$.

difficult to diagnose by looking at the prediction accuracy, Fig. 3(b). When the true energy is underestimated by 15%, the prediction accuracy is still 94% ($C = 10^2, N = 10^3$), which is only 1% lower than the highest value obtained with large N . Since we do not have any prior information about the maximum possible prediction accuracy for specific experimental datasets, these figures suggest that, judging by the prediction accuracy only, one can never be sure if the learned energy is a good estimate of the true one: a seemingly high accuracy is not enough.

V. PRESENCE OF REDUNDANT FEATURES IN THE INPUT DATA DEGRADES THE QUALITY OF THE INFERENCE

In Ref. [8], 166 inputs were used for predicting rearrangements. However, many of these inputs were correlated with one another. To model this, we repeat our analysis using a higher-dimensional input vector. For this, as explained in Sec. II, we train the SVM on a 20-dimensional input. Of these input dimensions, $x_i^j, j = 1, \dots, 10$ were independently sampled from a Gaussian distribution, and the remaining inputs were strongly nonlinearly correlated with them. We again train an SVM on $N = 10^6$ balanced data points. The

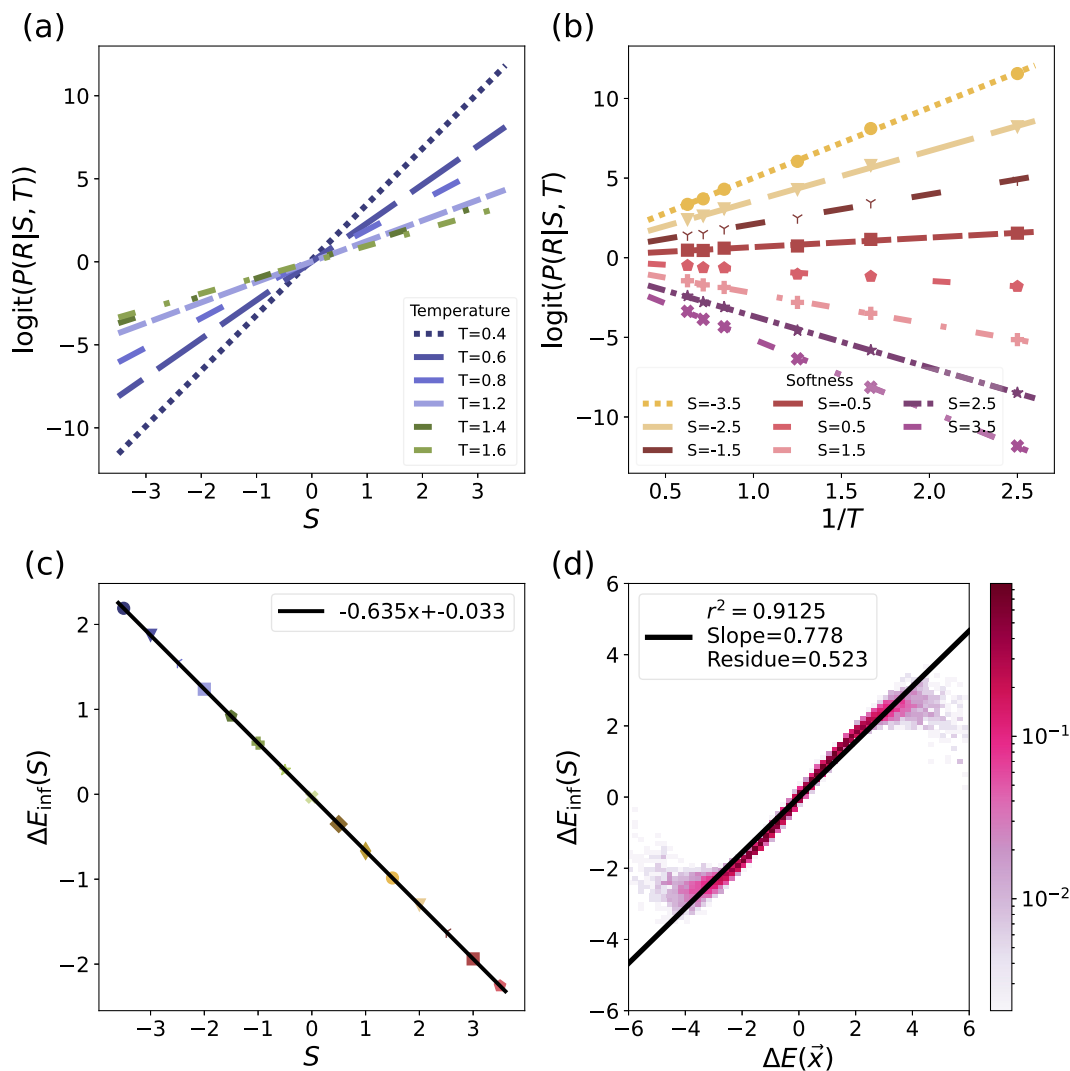


FIG. 4. Relationship between softness and $\Delta E(\vec{x})$ for a symmetric distribution of training energies and with spurious, correlated input terms. The same plotting conventions are used as in Fig. 1. In (d) the true energy barrier $\Delta E(\vec{x})$ vs the inferred energy barrier from SVM $\Delta E_{\text{inf}}(S)$ is plotted. The error to the fit is given by the purple semitransparent spread on both sides of the fit on a 2D density plot. Note that, to the extent that the slope in (d) is not 1, the correct energy is not learned. Also, the deviation between the fit and the 2D density plot at the edges shows that even though a linear fit was used to fit the energy and softness and its fit has a high r^2 value, the underlying function one is trying to fit is not really linear in S .

logit $P(R|S)$ vs S plot [Fig. 4(a)], logit $P(R|S, T)$ vs $1/T$ plot [Fig. 4(b)] and the inferred energy $\Delta E_{\text{inf}}(S)$ vs softness plot [Fig. 4(c)] again are linear, as in Fig. 1 and the previous work [8]. However, plotting the inferred energy $\Delta E_{\text{inf}}(S)$ vs the true energy barrier $\Delta E(\vec{x})$ for each configuration and producing a linear fit between them, cf. Fig. 4(d), we see that the magnitude of the inferred energy is underestimated compared to the true energy even for very large N (cf. Fig. 5). Looking at the optimal hyperplane learned by the SVM, we observe that the hyperplane contains contributions from the input variables that do not contribute to the true energy (not shown). One would not be aware of this problem from Figs. 4(a)–4(c) alone. We remind the reader that the true energy needed to produce Fig. 4(d) is typically unknown.

To design a method for identifying the bias from data, we note again that the variance of the true energy barrier distribution is a sum of the variance explained by S [i.e.,

the variance of $\langle \Delta E \rangle(S)$ over the distribution of S] and the variance conditional on S (i.e., the part of the energy barrier *not* captured by S). Thus, if we can find a different set of coordinates that allows the SVM to learn a different S that is *closer* to the true energy, this improvement should manifest as an increase in the variance of the distribution of inferred energies, $\text{Var}[\Delta E_{\text{inf}}]$. Our approach is then to reduce dimensionality of the input space, aiming to remove the correlated dimensions and increase the accuracy of the model at the same time. A particular version of this approach is known in the SVM literature as the recursive feature elimination (RFE) [40] procedure. RFE has been used in earlier work on predicting rearrangements [41,42] for pruning the dimensionality of SVM inputs. Assuming that all input dimensions are normalized to the same variance, RFE works by removing the input dimension with the smallest magnitude contribution to the separating hyperplane. One then refits the SVM and continues

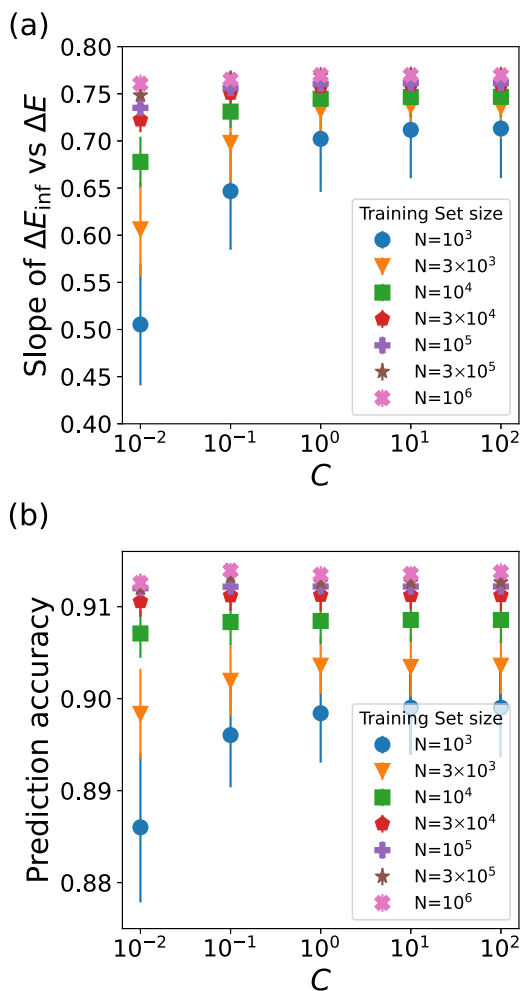


FIG. 5. (a) Slope of the inferred energy $\Delta E_{\text{inf}}(S)$ vs the true energy $\Delta E(\bar{x})$ and (b) the prediction accuracy for the model with spurious correlated inputs. Same plotting conventions as in Fig. 3.

the process iteratively. Figure 6(a) shows the variance of the inferred energy as a function of the number of inputs kept by the RFE procedure. The peak in $\text{Var}[\Delta E_{\text{inf}}]$ clearly matches the true number of dimensions that contribute to the energy in our model. Figure 6(b) shows a corresponding (but broader) peak in the prediction accuracy as well. These analyses bode well for using RFE for pruning the input data and resulting in a more accurate inference of the energy barrier in real-world problems.

VI. DISCUSSION

We have shown that, in our toy model, one can always use a linear SVM to predict rearrangements with a high accuracy, though the amount of data needed for this might be larger than what typical experiments would allow in realistic cases. However, even if the inference seems successful, the inferred energy barrier matches the true energy only in specific cases. Crucially, by observing a high prediction accuracy or high-quality linear relationship between softness, log rearrangement probability, and $1/T$, one cannot conclude that the correct energy has been learned. The problem

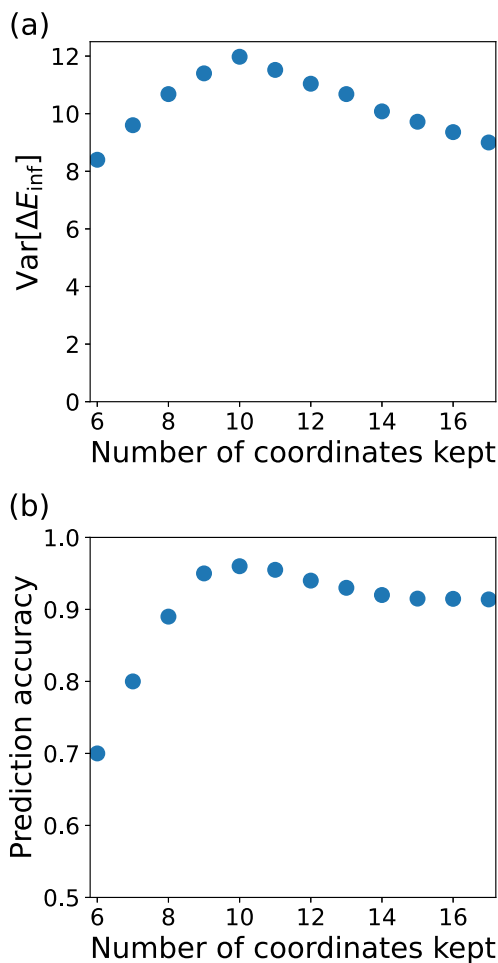


FIG. 6. Plot of variance of ΔE_{inf} and the prediction accuracy as a function of number of coordinates kept. The variance of the distribution of ΔE_{inf} is more sensitive for detecting relevant dimensions. The variance of inferred energy at the peak matches well with the variance of the distribution of true energy (12 in our units).

becomes severe—even in our simple model—when the input data has extra features, potentially nonlinearly correlated with true variables describing the model. Realistic systems, e.g., glasses, are likely to have different types of correlations between their input features than those we have considered here. Nonetheless, our results suggest a need to carefully scrutinize the use of ML methods, and specifically SVMs, for inference of energy barriers in glasses.

For our model, we have demonstrated a method to diagnose and fix this problem: RFE can be used to remove “confusing” input features. By tracking the variance of the inferred energy barriers or of $\log P(R|S)$, which is maximal when the true barriers are learned, improvements in the inference of the barriers can be detected, even though the true barriers are unknown and the prediction accuracy may change little. RFE is particularly natural in our problem because there is a clear division between important and unnecessary input dimensions. For other systems, RFE may not be the best method for adjusting the set of input features. For example, Appendix D 2 shows an example of a set of correlated features where no smaller set is sufficient to express the energy, and thus RFE cannot

recover the true energy. As another example, if the input features are a discretization of the pair correlation function $g(r)$, it may be more natural to coarsen this discretization, or to change the choice of basis functions, than to eliminate specific input features. However, our criterion for comparing different choices of input features in general would still stand: features that produce a larger variance in the inferred energy barriers should be closer to predicting the true barriers. We expect it to be true in general that the choice of features for the inference will affect whether or not the true energy is learned, so that different possible choices should be compared using this criterion. The need to make such comparisons between different choices of input features and other hyperparameters, rather than only focusing on achieving the best possible prediction accuracy, is one of the main conclusions of our work.

In our simple model, the probability of particle rearrangement is purely a function of energy. However, when SVMs are used to predict rearrangements in real systems, the probability is a function of energy as well as of an entropic prefactor, both of which are found to depend on S [8], see Eq. (1). In addition, there are other complications not present in our toy model, such as ambiguity in the identification of rearrangements. We expect such complications to only strengthen our conclusion that a good prediction accuracy does not guarantee that the ML model learns the true values of the energy barriers.

It may seem surprising that the addition of extra coordinates degrades the prediction accuracy and the quality of inference of ΔE_{inf} . Conventional wisdom is that such overcomplete representation should improve SVM accuracy by creating a higher-dimensional embedding space, in which the data become linearly separable [39]. It is possible that the failure of this intuition in our case comes from the probabilistic nature of rearrangements: for any \bar{x} , there are both rearranging and nonrearranging examples, at least in the $N \rightarrow \infty$ limit. Thus, the data are fundamentally not separable, irrespective of the space in which we embed them.

The process of adding more correlated coordinates explicitly to our input is similar to using some nonlinear kernel on the original data. SVM kernels allow us to create high-dimensional embeddings that are nonlinear functions of the input coordinates without having to explicitly evaluate the embedding, and these embeddings are often even infinite-dimensional. Thus, our results seem to imply that using a kernel may also prevent the true energy barriers from being learned.

In our work, we have focused specifically on linear SVMs, rather than other ML methods, because this is the only method which has been used in the past to explicitly deduce the underlying energy barriers from the inferred statistical model. However, note that we have chosen the true energy function to be expressible by a linear SVM. Further, note that more complex ML methods are generally thought to behave similarly to kernel methods [43,44]. Thus, we expect that our results are not caused by the simplicity of linear SVMs, and they will generalize to other ML approaches to the problem of learning energy barriers in glassy systems.

Our results may have implications for many systems beyond supercooled liquids, for which the underlying “physics” must be learned from an ML model trained on the data.

Indeed, we have shown that, even given a powerful ML model that can express the true underlying physics, an arbitrarily large amount of training data, and a good prediction accuracy, the model may fail to learn a correct physical description even in a relatively simple scenario. We suspect that, in real-world applications, this problem will become even more severe. One must then use independent methods—going beyond prediction accuracy—to evaluate the model quality.

ACKNOWLEDGMENTS

We thank Andrea Liu, Daniel Sussman, Eric Weeks, Daniel Weissman, and Ahmed Roman for important feedback. This work was funded, in part, by a Simons Foundation Investigator grant and NIH Grant No. 5-R01-NS084844. We also acknowledge the use of the HyPER C3 cluster of Emory University’s AI.Humanity Initiative.

APPENDIX A: THE INFERRED ENERGY AND PREDICTORS IN REAL GLASS SIMULATIONS CAN BE APPROXIMATED BY GAUSSIAN

We looked at the distribution of each of the 266 dimensions used to train the SVM for the Kob-Anderson model supercooled liquid [36]. All the dimensions looked unimodal. We calculated the kurtosis of all the dimensions, which measures how heavy tailed or light tailed a distribution is compared to a Gaussian distribution, which has kurtosis 0. A total of 73% of the dimensions had a kurtosis in the range of $(-0.3, 0.3)$ and 92% of the dimensions had a kurtosis in the range of $(-2, 2)$. The values of kurtosis cutoffs acceptable for normality vary widely from ± 2 to ± 6 [45–49], and thus the true structural features have roughly Gaussian distributions.

APPENDIX B: QUALITATIVE RESULTS REMAIN UNCHANGED WHEN TRAINED ON DATA WITH NONCENTERED DISTRIBUTION OF ENERGY BARRIERS

Recall, as explained in the main text, that in the true system all energy barriers are positive. However, in the main text, we chose energy barriers to be symmetric around zero for simplicity. Figure 7 is the analog of Fig. 1, but now evaluated for a model where almost all energy barriers are positive. We balance the training set, similarly to Ref. [8], so that the number of rearranging and nonrearranging particles is the same.

We draw each of the dimensions from a Gaussian distribution with unit variance centered at zero. We have ten independently sampled input dimensions, with $\alpha = 0.4$ for $j = 1, \dots, 10$. Further, we add a constant to the energy so that the mean of the distribution is two standard deviations away from zero, and thus the energy is almost always positive. We use $N = 3 \times 10^5$ training samples with 1.5×10^5 examples each of rearranging and nonrearranging configurations to train the SVM. As seen in Fig. 7, the results for the probability of rearrangement and the inferred energy remain qualitatively unchanged from Fig. 1 in the main text. In particular, the correct energy barriers are learned.

Just as in the case with a centered ΔE distribution, with a noncentered ΔE distribution, the energy is not learned

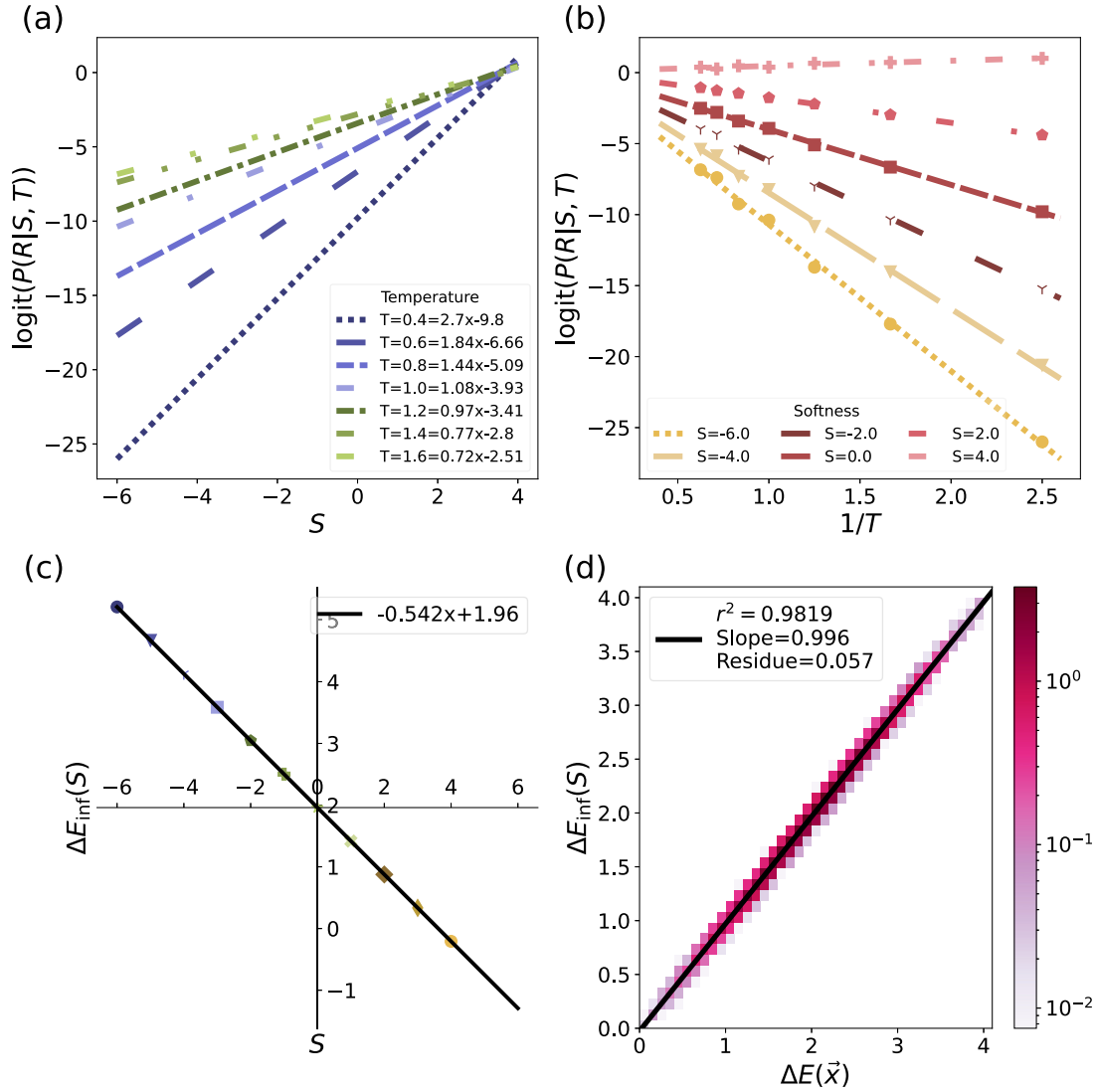


FIG. 7. Relationship between softness and $\Delta E(\vec{x})$ for positive energy barriers and a balanced dataset. To balance our dataset, we choose 50% of samples where rearrangement was observed, and 50% where it was not. Plotting conventions are the same as in Fig. 1. Note that the correct energy is learned (slope of 0.996), and the spread in the 2D density plot is minimal.

correctly at small training sample sizes. We generated a non-centered ΔE distribution as above, and generated training sets of different sizes, balancing them as above. As can be seen from Fig. 8, these observations are qualitatively the same as in the centered case.

We also note that giving each variable x^j a nonzero mean μ_j has no effect except producing a nonzero mean ΔE , and is thus expected to be covered by the above checks. To see this, write $x^j = \mu_j + y^j$, where y^j has mean zero. We then have

$$\Delta E = \sum_j \alpha^j \mu_j + \sum_j \alpha^j y^j. \quad (\text{B1})$$

Thus, the only effect of giving x^j nonzero mean is to add a constant $\sum \alpha^j \mu_j$ to ΔE . Further, note that even in this case where $\mu_j \neq 0$, changing the sign of α^j only changes the mean ΔE : it has no effect on $\sum \alpha^j y^j$, since the distribution of y is symmetric around 0. Thus, qualitative results such as the above, which hold both when the mean ΔE is 0 and when it

is positive, are expected to still hold when some of the α^j are negative.

APPENDIX C: EFFECT OF MISSING FEATURES

In any realistic system, some of the features needed to express ΔE will be missing. Here, we confirm that this prevents the correct energy from being learned. We use $N = 10^6$ training samples with 5×10^5 examples each of rearranging and nonrearranging configurations to train the SVM. The distribution of energies in the training sample is symmetric. We use ten independently sampled input dimensions, with $\alpha^j = 1.2$ for $j = 1, \dots, 10$, to determine the energy. Out of the ten dimensions we train the SVM only with the first nine dimensions and drop the last dimension. In this case, one ends up underestimating the variance of the true energy, as can be seen from Fig. 9.

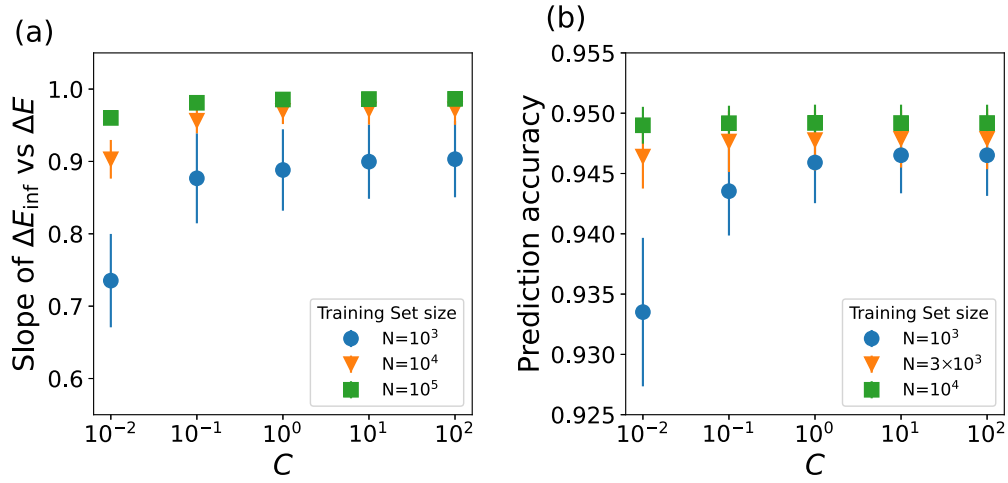


FIG. 8. (a) Slope of the inferred energy $\Delta E_{\text{inf}}(S)$ vs the true energy $\Delta E(\vec{x})$ and (b) the prediction accuracy for the model with noncentered ΔE distribution. Same plotting conventions as in Fig. 3.

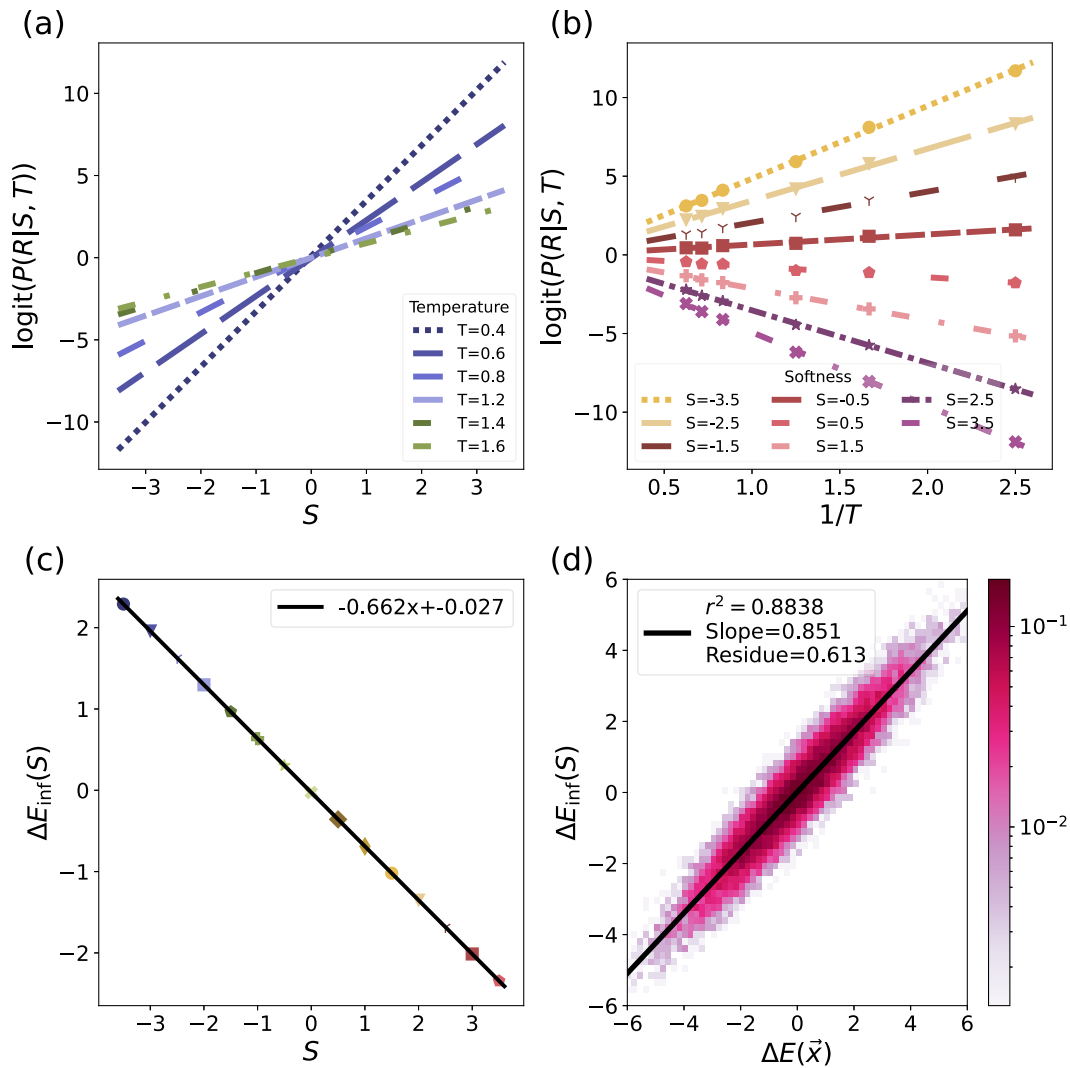


FIG. 9. Relationship between softness and $\Delta E(\vec{x})$ for symmetric distribution of training energies for a large training set size, $N = 10^6$, with one of the relevant feature missing. (a)–(c) Same as in Fig. 1. (d) The true energy barrier $\Delta E(\vec{x})$ vs the inferred energy barrier from SVM $\Delta E_{\text{inf}}(S)$. Note that, to the extent that the slope in (d) is not 1, the correct energy is not learned.

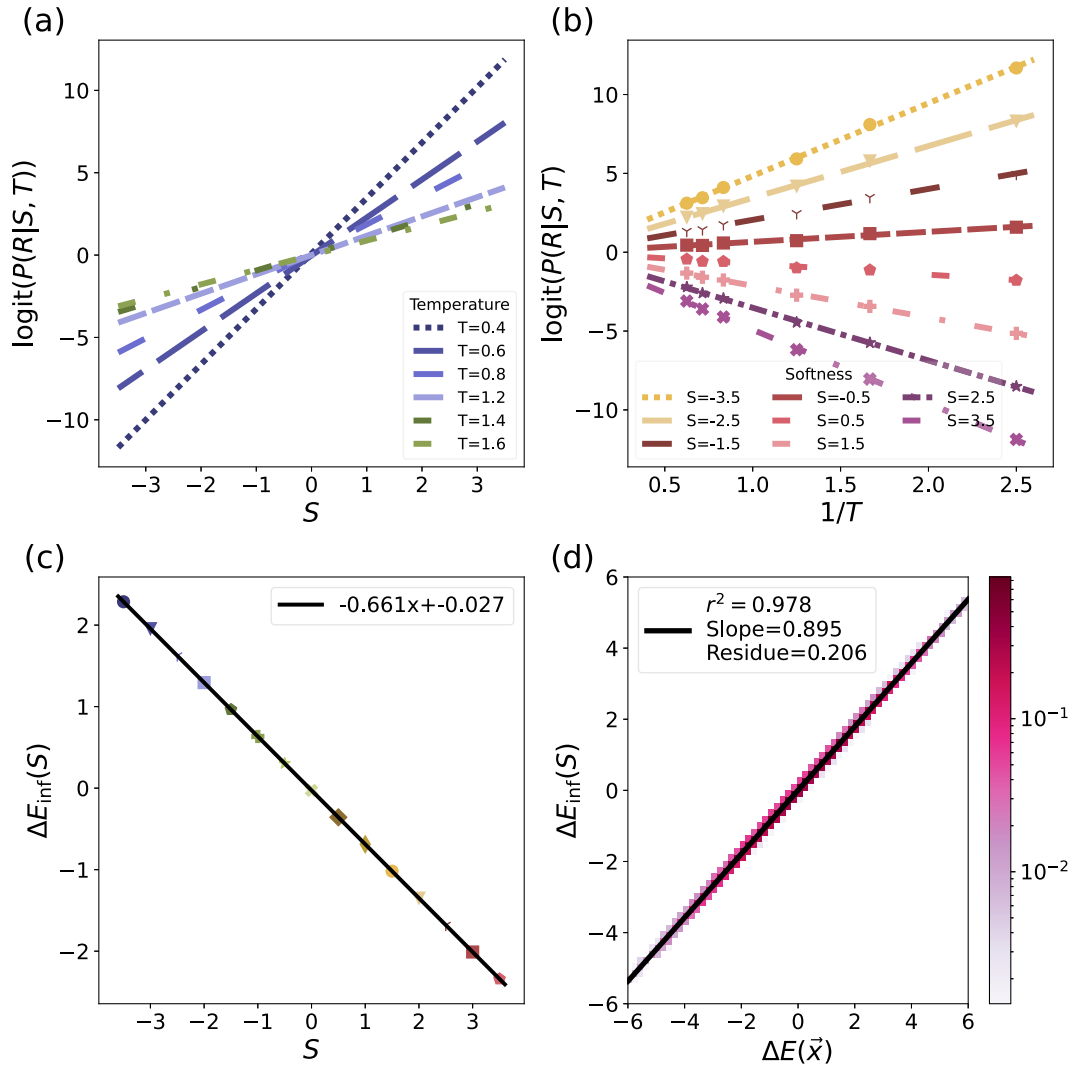


FIG. 10. Relationship between softness and $\Delta E(\vec{x})$ for symmetric distribution of training energies for a large training set size, $N = 10^6$, with additional linear features. (a)–(c) Same as in Fig. 1. (d) The true energy barrier $\Delta E(\vec{x})$ vs the inferred energy barrier from SVM $\Delta E_{\text{inf}}(S)$. Note that, to the extent that the slope in (d) is not 1, the correct energy is not learned.

APPENDIX D: EFFECT OF DIFFERENT CHOICES OF CORRELATED FEATURES

As our model of correlated features in Sec. V, we have chosen to add nonlinear functions of the “correct” input features to the input. Here, we check that the results of Sec. V generalize to other choices of correlated input features. In particular, we test two other options. Firstly, we consider addition of variables that, rather than being nonlinear functions of the “correct” input features, are simply linearly correlated with them. Secondly, we consider a set of input features that are nonlinearly correlated, but are not “redundant,” in the sense that, in principle, all of them are required to express the true energy through a linear function. In both cases, we find that the results of Sec. V remain qualitatively unchanged.

1. Effect of redundant linear feature

We use $N = 10^6$ training samples with 5×10^5 examples each of rearranging and nonrearranging configurations to train

the SVM. The distribution of energies in the training sample is symmetric. We have ten independently sampled input dimensions, with $\alpha^j = 1.2$ for $j = 1, \dots, 10$. Thus, ten dimensions determine the energy. We train the SVM on a 12-dimensional input which consists of all ten dimensions and one extra copy each of $j = 1, 2$. This gives two extra, redundant features which are linear in the relevant coordinates. In this case, the SVM again underestimates the variance of the true energy, as can be seen from Fig. 10.

2. Effect of nonredundant and nonlinear correlated features

We use $N = 10^6$ training samples with 5×10^5 examples each of rearranging and nonrearranging configurations to train the SVM. The distribution of energies in the training sample is symmetric. We have ten independently sampled input dimensions, with $\alpha^j = 1.2$ for $j = 1, \dots, 10$. Thus, ten dimensions determine the energy. Instead of giving the SVM $x^1 \dots x^{10}$ as

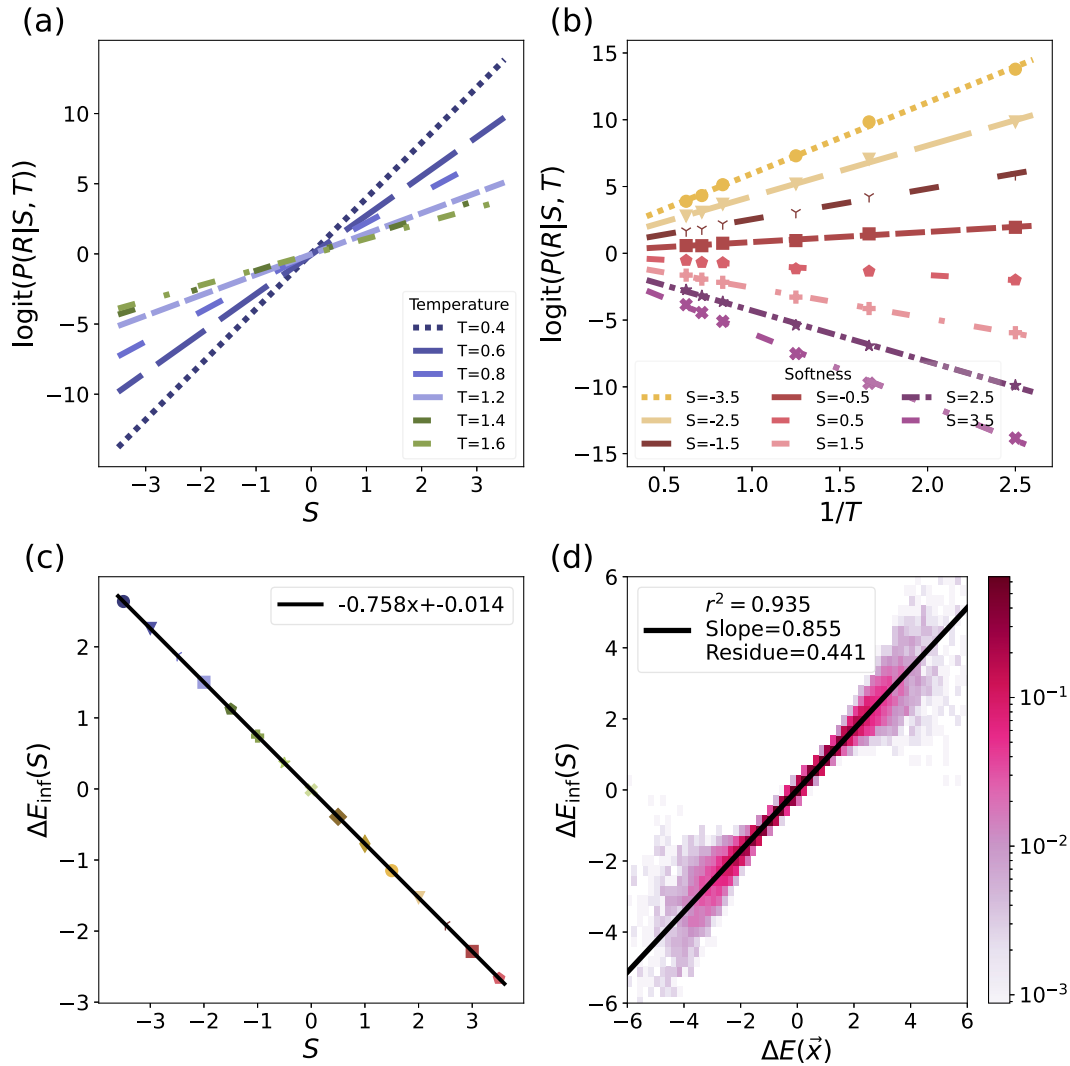


FIG. 11. Relationship between softness and $\Delta E(\vec{x})$ for a symmetric distribution of training energies and with spurious, correlated input terms. The same plotting conventions are used as in Fig. 1. (d) The true energy barrier $\Delta E(\vec{x})$ vs the inferred energy barrier from SVM $\Delta E_{\text{inf}}(S)$. The fit error is illustrated by the purple semitransparent spread on both sides of the fit on the 2D density plot. Note that, to the extent that the slope in (d) is not 1, the correct energy is not learned. Also, the deviation between the fit and the 2D density plot at the edges shows that, even though a linear fit was used to fit the energy and softness, and the fit had a high r^2 value, the underlying function we are trying to fit here is not linear in S .

input features, we use the 14 input features

$$x^1 + (x^5)^3, x^2 + (x^6)^3, x^3 + (x^7)^3, x^4 + (x^8)^3, x^5, x^6, \dots, x^{10}, (x^5)^3, (x^6)^3, (x^7)^3, (x^8)^3. \quad (D1)$$

(There is nothing particular about these features, and we believe that other combinations of powers of predictors would

deliver a similar point.) It should be possible for the SVM to learn a linear combination of these features that would cancel out the cubic terms and infer the true energy. Nonetheless, we observed that the SVM *does not* learn the correct energy, Fig. 11. Thus, the presence of nonlinearities as well as redundant features affects the ability of SVM to predict the correct energy.

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
 [2] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).

[3] W. Hu, R. R. P. Singh, and R. T. Scalettar, Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination, *Phys. Rev. E* **95**, 062122 (2017).
 [4] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, *Nat. Phys.* **13**, 435 (2017).

- [5] L. Wang, Discovering phase transitions with unsupervised learning, *Phys. Rev. B* **94**, 195105 (2016).
- [6] R. Wang, Y.-G. Ma, R. Wada, L.-W. Chen, W.-B. He, H.-L. Liu, and K.-J. Sun, Nuclear liquid-gas phase transition with machine learning, *Phys. Rev. Res.* **2**, 043202 (2020).
- [7] X. Zhao and L. Fu, Machine learning phase transition: An iterative proposal, *Ann. Phys.* **410**, 167938 (2019).
- [8] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, A structural approach to relaxation in glassy liquids, *Nat. Phys.* **12**, 469 (2016).
- [9] S. S. Schoenholz, E. D. Cubuk, E. Kaxiras, and A. J. Liu, Relationship between local structure and relaxation in out-of-equilibrium glassy systems, *Proc. Natl. Acad. Sci.* **114**, 263 (2017).
- [10] E. D. Cubuk *et al.*, Structure-property relationships from universal signatures of plasticity in disordered solids, *Science* **358**, 1033 (2017).
- [11] G. Biroli, Machine learning glasses, *Nat. Phys.* **16**, 373 (2020).
- [12] J. Greitemann, K. Liu, L. D. C. Jaubert, H. Yan, N. Shannon, and L. Pollet, Identification of emergent constraints and hidden order in frustrated magnets using tensorial kernel methods of machine learning, *Phys. Rev. B* **100**, 174408 (2019).
- [13] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, Identifying structural flow defects in disordered solids using machine-learning methods, *Phys. Rev. Lett.* **114**, 108001 (2015).
- [14] E. D. Cubuk, A. J. Liu, E. Kaxiras, and S. S. Schoenholz, Unifying framework for strong and fragile liquids via machine learning: A study of liquid silica, [arXiv:2008.09681](https://arxiv.org/abs/2008.09681) [cond-mat.soft].
- [15] V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. R. Nelson, T. Back, D. Hassabis, and P. Kohli, Unveiling the predictive power of static structure in glassy systems, *Nat. Phys.* **16**, 448 (2020).
- [16] E. Boattini, F. Smallenburg, and L. Filion, Averaging local structure to predict the dynamic propensity in supercooled liquids, *Phys. Rev. Lett.* **127**, 088007 (2021).
- [17] C. Giannetti, B. Lucini, and D. VDACCHINO, Machine learning as a universal tool for quantitative investigations of phase transitions, *Nucl. Phys. B* **944**, 114639 (2019).
- [18] D. M. Sussman, S. S. Schoenholz, E. D. Cubuk, and A. J. Liu, Disconnecting structure and dynamics in glassy thin films, *Proc. Natl. Acad. Sci.* **114**, 10601 (2017).
- [19] T. M. Obadiya and D. M. Sussman, Using fluid structures to encode predictions of glassy dynamics, *Phys. Rev. Res.* **5**, 043112 (2023).
- [20] N. Sun, J. Yi, P. Zhang, H. Shen, and H. Zhai, Deep learning topological invariants of band insulators, *Phys. Rev. B* **98**, 085402 (2018).
- [21] L. Berthier and G. Biroli, Glasses and aging, a statistical mechanics perspective on, in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers (Springer, New York, 2009), pp. 4209–4240.
- [22] M. D. Ediger, Spatially heterogeneous dynamics in supercooled liquids, *Annu. Rev. Phys. Chem.* **51**, 99 (2000).
- [23] W. Kob, C. Donati, S. J. Plimpton, P. H. Poole, and S. C. Glotzer, Dynamical heterogeneities in a supercooled Lennard-Jones liquid, *Phys. Rev. Lett.* **79**, 2827 (1997).
- [24] I. Tah and S. Karmakar, Signature of dynamical heterogeneity in spatial correlations of particle displacement and its temporal evolution in supercooled liquids, *Phys. Rev. Res.* **2**, 022067(R) (2020).
- [25] J. C. Mauro, Grand challenges in glass science, *Front. Mater.* **1**, 20 (2014).
- [26] G. Biroli and J. P. Garrahan, Perspective: The glass transition, *J. Chem. Phys.* **138**, 12A301 (2013).
- [27] I. Tah, S. A. Ridout, and A. J. Liu, Fragility in glassy liquids: A structural approach based on machine learning, *J. Chem. Phys.* **157**, 124501 (2022).
- [28] G. Jung, G. Biroli, and L. Berthier, Predicting dynamic heterogeneity in glass-forming liquids by physics-inspired machine learning, *Phys. Rev. Lett.* **130**, 238202 (2023).
- [29] G. Jung, G. Biroli, and L. Berthier, Dynamic heterogeneity at the experimental glass transition predicted by transferable machine learning, *Phys. Rev. B* **109**, 064205 (2024).
- [30] F. S. Pezzicoli, G. Charpiat, and F. P. Landes, Rotation-equivariant graph neural networks for learning glassy liquids representations, *SciPost Phys.* **16**, 136 (2024).
- [31] X. Jiang, Z. Tian, K. Li, and W. Hu, A geometry-enhanced graph neural network for learning the smoothness of glassy dynamics from static structure, *J. Chem. Phys.* **159**, 144504 (2023).
- [32] R. M. Alkemade, E. Boattini, L. Filion, and F. Smallenburg, Comparing machine learning techniques for predicting glassy dynamics, *J. Chem. Phys.* **156**, 204503 (2022).
- [33] R. M. Alkemade, F. Smallenburg, and L. Filion, Improving the prediction of glassy dynamics by pinpointing the local cage, *J. Chem. Phys.* **158**, 134512 (2023).
- [34] G. Zhang, H. Xiao, E. Yang, R. J. S. Ivancic, S. A. Ridout, R. A. Riggleman, D. J. Durian, and A. J. Liu, Structuro-elastoplasticity model for large deformation of disordered solids, *Phys. Rev. Res.* **4**, 043026 (2022).
- [35] H. Xiao, G. Zhang, E. Yang, R. Ivancic, S. Ridout, R. Riggleman, D. J. Durian, and A. J. Liu, Identifying microscopic factors that influence ductility in disordered solids, *Proc. Natl. Acad. Sci.* **120**, e2307552120 (2023).
- [36] S. A. Ridout, I. Tah, and A. J. Liu, Building a “trap model” of glassy dynamics from a local structural predictor of rearrangements, *Europhys. Lett.* **144**, 47001 (2023).
- [37] S. A. Ridout and A. J. Liu, The dynamics of machine-learned “softness” in supercooled liquids describe dynamical heterogeneity, [arXiv:2406.05868](https://arxiv.org/abs/2406.05868) [cond-mat.soft].
- [38] B. E. Boser, I. M. Guyon, and V. N. Vapnik, A training algorithm for optimal margin classifiers, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92* (Association for Computing Machinery, New York, NY, 1992), pp. 144–152; C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning* **20**, 273 (1995).
- [39] B. Scholkopf, K.-K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, Comparing support vector machines with Gaussian kernels to radial basis function classifiers, *IEEE Trans. Signal Process.* **45**, 2758 (1997).
- [40] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, *Mach. Learn.* **46**, 389 (2002).
- [41] M. Harrington, A. J. Liu, and D. J. Durian, Machine learning characterization of structural defects in amorphous packings of dimers and ellipses, *Phys. Rev. E* **99**, 022903 (2019).

- [42] R. J. S. Ivancic and R. A. Riggleman, Identifying structural signatures of shear banding in model polymer nanopillars, *Soft Matter* **15**, 4548 (2019).
- [43] A. Jacot, F. Gabriel, and C. Hongler, Neural tangent kernel: Convergence and generalization in neural networks, *Adv. Neural Info. Proc. Sys.* **2018**, 8571 (2018).
- [44] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory* (Cambridge University, Cambridge, 2022).
- [45] N. Heckert, J. Filliben, C. Croarkin, B. Hembree, W. Guthrie, P. Tobias, and J. Prinz, *Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods* (National Institute of Standards and Technology, Gaithersburg, 2002).
- [46] D. George and P. Mallery, *SPSS for Windows step by step: A simple guide and reference, 17.0 update*, 10th ed. (Allyn & Bacon, Boston, 2010).
- [47] J. Hair, W. Black, and B. Babin, *Multivariate Data Analysis: A Global Perspective*, Global Edition (Pearson Education, London, 2010).
- [48] B. M. Byrne, *Structural Equation Modeling with Mplus, Basic Concepts, Applications, and Programming* (Routledge, New York, 2013).
- [49] T. K. Burdenski, Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures, *Multiple Linear Regression Viewpoints* **26**, 15 (2000).