# Policy Learning in High Dimensional Settings

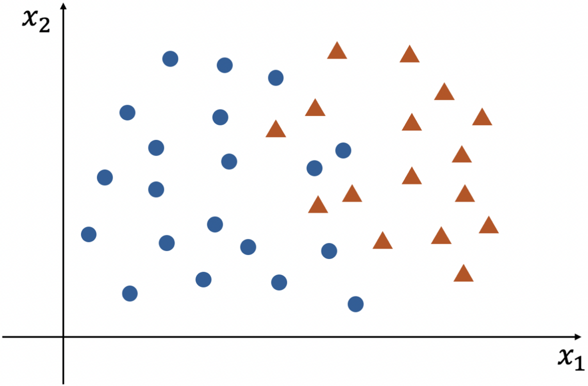
Presentation by: Eddie (Yidi) Wu

Brown University

April 13, 2026

# Overview

# Classification Problem




- A sample $\{X_i, t_i\}_{i=1}^{n}$ drawn from distribution $P$ over $\mathcal{X} \times \{+1, -1\}$.
- Want to learn a function $g$ that maps from $\mathcal{X}$ to $\{+1, -1\}$ to minimize risk.

# Population and Empirical Risk

- Population risk or classification risk of a function $g$:

$$R(f) = E_P[\mathbf{1}\{t_i g(\mathbf{x}) < 0\}]$$

- Empirical risk or sample risk of a function $g$:

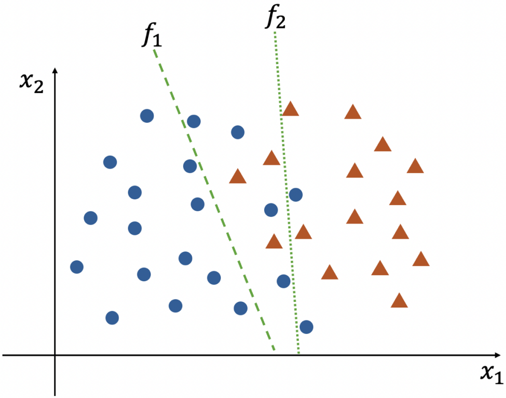
$$\hat{R}(f) = \frac{1}{m}\sum_{i=1}^{m}\mathbf{1}\{t_i g(\mathbf{x_i}) < 0\}$$

- Every binary classifier $g$ can be written as:

$$g = sign(f(\mathbf{x_i}))$$

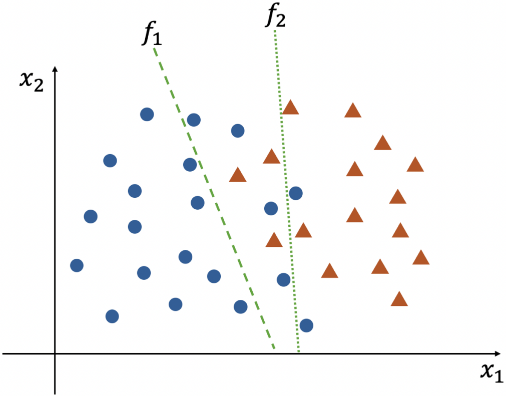
  where $f(\mathbf{x_i})$ is a real valued function.
- Examples of $f$ can be linear function, neural network etc.

# How to Learn $f$?




- Empirical risk minimization (ERM) principle: fix a hypothesis class e.g. the class of all linear functions.
- Choose the function in the hypothesis class that minimizes empirical risk.

# How to Learn $f$?




- 20 circle, 16 triangles in the sample.
- $f_1$ has risk $\frac{6}{36}$, $f_2$ has risk $\frac{5}{36}$. ERM would pick $f_2$ over $f_1$.

# Surrogate Loss and Optimization

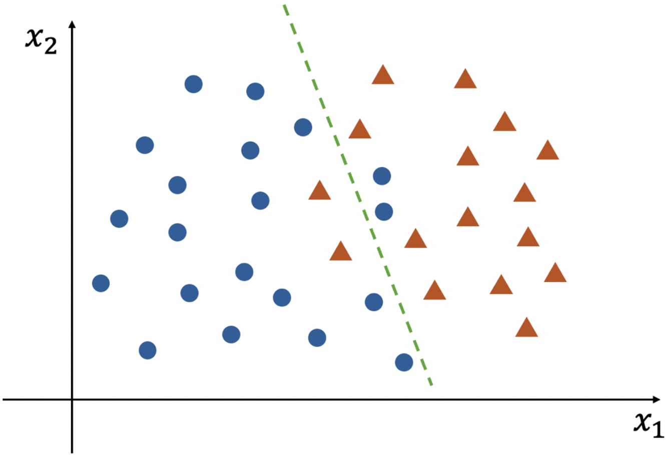
- Fixing a hypothesis class, minimizing $\hat{R}(f)$ directly is not tractable most of the times.
- Minimize surrogate risk instead:

$$\hat{R}_\phi(f) = \frac{1}{m}\sum_{i=1}^{m}\phi(t_i f(\mathbf{x_i}))$$

- Examples of $\phi$ can be $\phi(\alpha) = \max\{0, 1-\alpha\}$.
- Numerical optimization such as gradient descent.
- Please refer to Kitagawa et al (2023) for a comprehensive treatment of this topic.

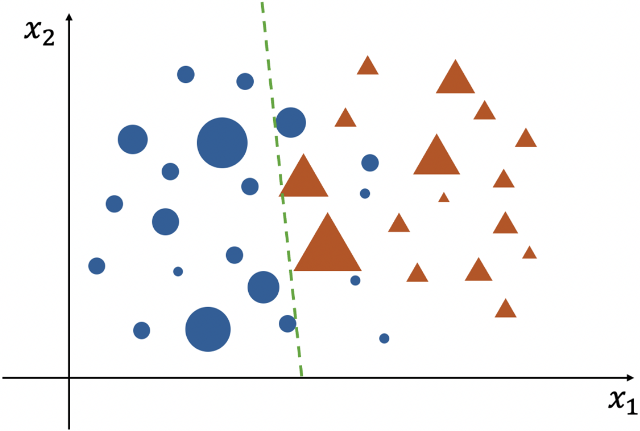
# Overview

# Unweighted Classification



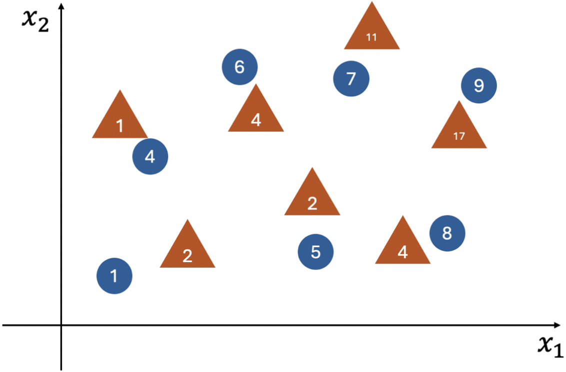

- Decision boundary for an unweighted classification problem.

# Weighted Classification



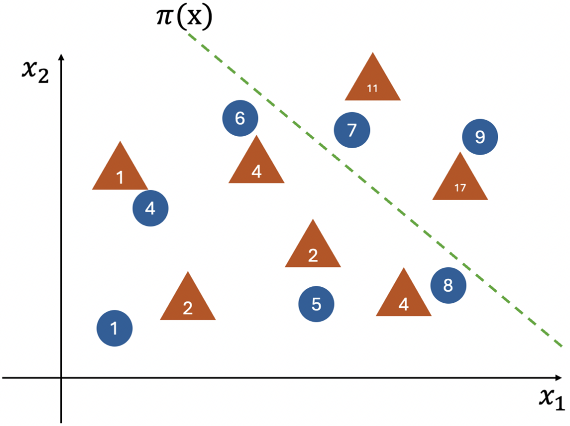

- Decision boundary for an weighted classification problem.
- Weighted empirical risk: $\hat{R}^{\omega}(f) = \frac{1}{m}\sum_{i=1}^{m} \omega_i \mathbf{1}\{t_i sign(f(\mathbf{x_i})) < 0)\}$.

# Policy Learning




- A sample $\{X_i, D_i, Y_i\}_{i=1}^m$ drawn from distribution $Q$.
- Want to learn a function $\pi$ that maps from $\mathcal{X}$ to $\{+1, -1\}$ to maximize population welfare.
  $W(f) := E_Q[y_i(+1)\mathbf{1}\{f(\mathbf{x_i}) \geq 0\} + y_i(-1)\mathbf{1}\{f(\mathbf{x_i}) < 0\}]$

# Policy Learning Intuition




- Example of a treatment assignment rule.
- Oracle policy is to assign treatment according to $CATE(\mathbf{x_i}) \geq 0$.

# Classification and Policy Learning

## Proposition (Risk minimization and welfare maximization (Kitagawa et al (2023)))

*Assuming $e(x)$ and $Q$ satisfy the following properties:*

- *Overlap: propensity score $0 < e(x) < 1$ for all $x \in \mathcal{X}$*
- *Unconfoundedness: $y_i(+1), y_i(-1) \perp D_i | \mathbf{x_i}$*

*Then the maximization of additive welfare criterion $W(f)$ is equivalent to the minimization of weighted classification risk*

$$E_Q[\omega_i \mathbf{1}\{t_i sign(f(\mathbf{x_i})) < 0\}]$$

*where $t_i = sign(y_i)D_i$ and $\omega_i = \frac{|y_i|}{D_i e(\mathbf{x_i}) + (1 - D_i)/2}$.*

- $W(f) := E_Q[y_i(+1)\mathbf{1}\{f(\mathbf{x_i}) \geq 0\} + y_i(-1)\mathbf{1}\{f(\mathbf{x_i}) < 0\}]$

# Classification and Policy Learning

- If a unit has large observed outcome, then it is valuable for the model to match the treatment status $D_i$ of the unit because mismatching $sign(f(\mathbf{x_i}))$ and $D_i$ is costly.
- Population welfare in terms of conditional average treatment effect $\tau(\mathbf{x_i})$:

$$W(f) = E_Q[y_i(-1)] + E_Q[\mathbf{1}\{sign(f(\mathbf{x_i})) \geq 0\}\tau(\mathbf{x_i})]$$

- The weighted classification boundary is in fact the CATE threshold at 0.
- What happens if we conduct policy learning with high dimensional features but limited sample size?

# Overview

# Contributions

- Demonstrated that by doing policy learning via weighted classification, “double ascent” phenomenon of welfare exists in overparameterized / high dimensional settings.
- Using a rich set of features and highly overparameterized models might lead to higher welfare than using a restricted model.

# Contributions

- Extended existing results on generalization bounds to weighted classification and additive welfare maximization.
- Explained the “double ascent” phenomena via the above generalization bounds by building on Lee & Cherkassky (2024).
- Suggestions for practitioners: potential value of flexible, high dimensional policy rules using rich covariate information.

# Data Generating Process

We use the following RCT data generation set-up. $n$ units $\{\mathbf{x_i}, D_i, y_i(+1), y_i(-1)\}_{i=1}^m$ are randomly drawn from:

$$\mathbf{x_i} \sim N_d(0, \sigma_x^2 I_d) \tag{1}$$

$$y_i(+1) = \mathbf{x_i}^T \beta_1 + \epsilon_{1,i} \tag{2}$$

$$y_i(-1) = \mathbf{x_i}^T \beta_0 + \epsilon_{0,i} \tag{3}$$

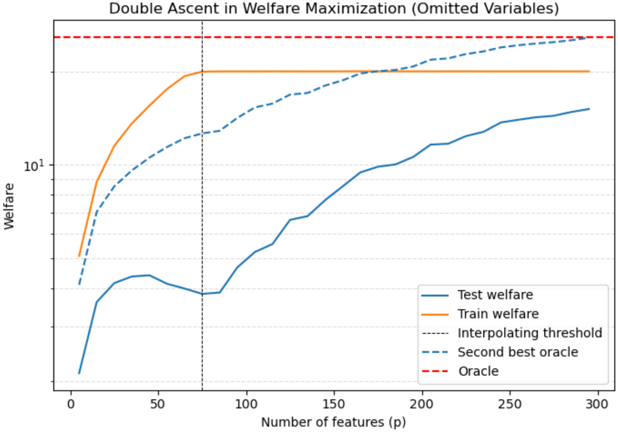
$$D_i \sim \textit{symmetric Bernoulli}(q) \tag{4}$$

- Potential outcomes are not observed. We observe a sample $\{\mathbf{x_i}, D_i, y_i\}_{i=1}^m$, where $y_i = (\frac{1+D_i}{2}) \cdot y_i(+1) + (\frac{1-D_i}{2}) \cdot y_i(-1)$.
- $CATE(\mathbf{x_i}) = \mathbf{x_i}^T(\beta_1 - \beta_0)$.
- $\epsilon_{1,i}$ and $\epsilon_{0,i}$ have zero mean.

# Omitted Variables

- Sample size $m$ fixed at 150. $\mathbf{x_i} \in \mathbb{R}^{300}$.
- For $p \in [5, 10, ..300]$:
  - Take the first $p$ features of $\mathbf{x_i}$, denote it by $\tilde{x}_{ip}$.
  - Fit $g(\tilde{x}_{ip}) = sign(\tilde{x}_{ip}^T \alpha_p)$ by optimizing $\mathcal{L}(\alpha_p) = \frac{1}{m} \sum_{i=1}^{m} \omega_i \phi(t_i \tilde{x}_{ip}^T \alpha_p)$ using gradient descent, where $\phi(c) = \log(1 + e^{-c})$ logistic loss.
  - Signal strength increases linearly in $\beta_1$ and $\beta_0$, though this can be relaxed and obtain qualitatively similar results.
- Repeat the above and average over 200 runs.

# Omitted Variables results

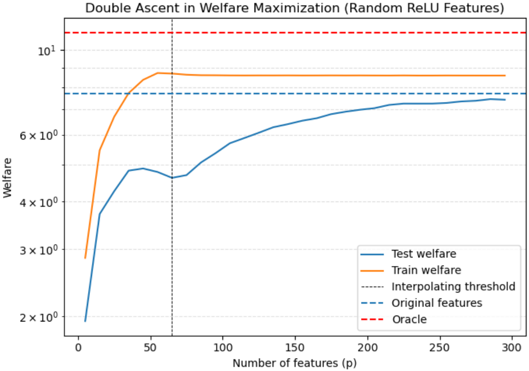



- Gap between the two oracles due to omitted variables.
- Gap between second best oracle and test welfare due to sampling error and using surrogate loss.
- Second ascent in welfare despite heavily overparameterized.

# Random Features

- Sample size $m$ fixed at 150. $\mathbf{x_i} \in \mathbb{R}^{50}$.
- Instead of observing $\mathbf{x_i}$, for $p \in [5, 10, ..300]$,:
  - We observe $\mathbf{z_i} = \sigma(W_p \mathbf{x_i})$ where $W_p \in \mathbb{R}^{p \times 50}$ is a random matrix drawn from multivariate normal. $\sigma(.)$ is a nonlinear activation function.
  - Fit $g(\mathbf{z_i}) = sign(\mathbf{z_i}^T \alpha_p)$ by optimizing $\mathcal{L}(\alpha_p) = \frac{1}{m} \sum_{i=1}^{m} \omega_i \phi(t_i \mathbf{z_i}^T \alpha_p)$ using gradient descent, where $\phi(c) = \log(1 + e^{-c})$ logistic loss.
- Repeat and average over 200 runs.
- Akin to observing noisy measures of a set of lower dimensional true features.

# Random Feature Results




- Gap between oracle and using original features due to sampling error.
- Gap between original features and test welfare due to surrogate loss and parameterization.
- Regardless of the dimension of $\mathbf{x_i}$, test welfare converges to original features as $p/n$ increases.

# Other Simulations

"Double ascent" is also observed when:

- True CATE is nonlinear in $\mathbf{x_i}$ but we fit the same linear treatment assignment rule.
- True CATE is linear in $\mathbf{x_i}$ and for each value of $p$, the true DGP has the same feature dimensions as the observed sample, i.e. "moving target".

# Overview

# Quantifying Complexities

- Empirical Rademacher Complexity of a set of bounded real valued functions $\mathcal{G}$, fixing a sample $S$:

$$\hat{\mathfrak{R}}_S(\mathcal{G}) := E_\sigma \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i g(x_i) \right] = E_\sigma \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sigma \cdot \mathbf{g} \right]$$

- How well the functions in $\mathcal{G}$ can fit random noise on average.

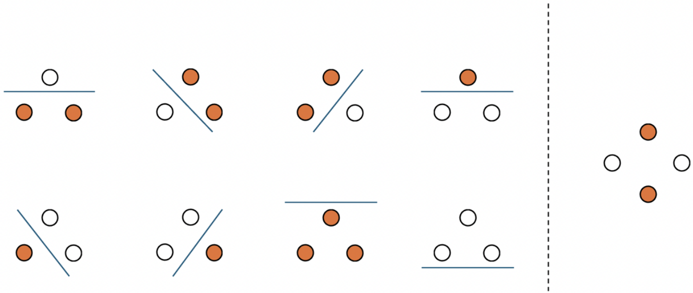
# Quantifying Complexities

- VC dimension of a set of binary functions $\mathcal{G}$ is the size of the largest set that can be shattered by $\mathcal{G}$:

$$VCdim(\mathcal{G}) = \max\{m : \Pi_{\mathcal{G}}(m) = 2^m\}$$

where $\Pi_{\mathcal{G}}(m)$ is the maximum number of distinct ways in which $m$ points can be classified using hypothesis in $\mathcal{G}$.

# VC Dimension example



The VC dimension of a hyperplane in $\mathbb{R}^n$ is $n + 1$.

# Bounded Weights

## Proposition (VC-dimension generalization bound for weighted classification)

*Let $\mathcal{H}$ be a class of functions taking values in $\{+1, -1\}$ with VC-dimension d. Assume that the classification weight $\omega$ is bounded between $0 \leq \omega \leq W$ where $W = \sup \omega$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over a sample S of size m, the following holds for all $f \in \mathcal{H}$:*

$$R^{\omega}(f) \leq \hat{R}^{\omega}(f) + W\left(\sqrt{\frac{2d \log \frac{em}{d}}{m}} + \sqrt{\frac{\log \frac{1}{\delta}}{2m}}\right) \tag{5}$$

$$= O\left(W\sqrt{\frac{\log \frac{m}{d}}{\frac{m}{d}}}\right) \tag{6}$$

*where m is the size of the training sample.*

# Proof Sketch

- Uniformly bound the empirical process $R^{\omega}(f) - \hat{R}^{\omega}(f)$ using McDiarmid's Inequality in terms of Rademacher complexity.
- Then bound Rademacher complexity with VC dimension.

# Sub-gaussian Weights

## Proposition (VC-dimension generalization bound with sub-Gaussian weights)

*Let $\mathcal{H}$ be a class of functions taking values in $\{+1, -1\}$ with VC-dimension d. Assume that the classification weight $\omega$ is sub-Gaussian with constant $K$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over a sample $S$ of size m, the following holds for all $f \in \mathcal{H}$:*

$$R^{\omega}(f) \leq \hat{R}^{\omega}(f) + K\sqrt{\log \frac{2m}{\delta}} \left( 2\sqrt{\frac{d \log \frac{em}{d}}{m}} + \sqrt{\frac{\log \frac{2}{\delta}}{m}} \right) \tag{7}$$

$$= O\left( \sqrt{\log m} \sqrt{\frac{\log \frac{m}{d}}{\frac{m}{d}}} \right) \tag{8}$$

*where m is the size of the training sample.*

# Welfare Generalization Bound

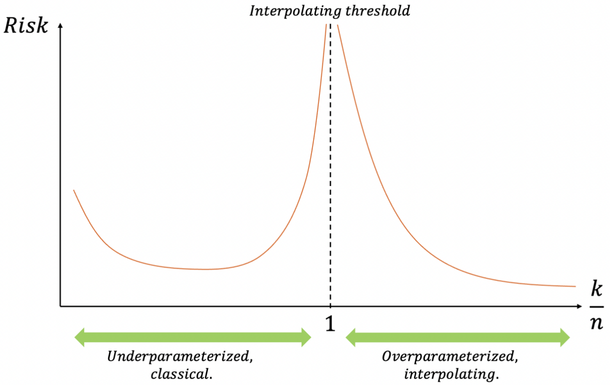
## Corollary (VC-dimension generalization bound for welfare)

*Let $\mathcal{H}$ be a class of functions taking values in $\{+1, -1\}$ with VC-dimension d. Suppose that the population distribution is Q, and that the classification weight, which is defined as $\omega := \frac{|y_i|}{D_i e(\mathbf{x_i}) + (1-D_i)/2}$, is sub-Gaussian with constant K. Then, for any $\delta > 0$, with probability at least $1-\delta$ over a sample S of size m, the following holds for all $f \in \mathcal{H}$:*

$$W(f) \geq \hat{W}(f) - K\sqrt{\log\frac{2m}{\delta}}\left(2\sqrt{\frac{d\log\frac{em}{d}}{m}} + \sqrt{\frac{\log\frac{2}{\delta}}{m}}\right) + C_{Q,m} \quad (9)$$

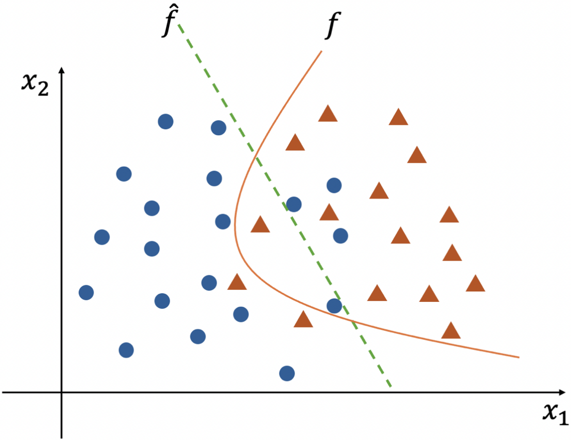
*where m is the size of the training sample, $C_{Q,m}$ is a constant that depends only the population distribution Q and the random sample and is independent of f, and empirical welfare as a function of treatment assignment rule f is defined as*
$\hat{W}(f) = \frac{1}{n}\sum_{i=1}^{n}(y_i(+1)\mathbf{1}\{f(\mathbf{x_i}) \geq 0\} + y_i(-1)\mathbf{1}\{f(\mathbf{x_i}) < 0\})$.

# Underparameterized Regime



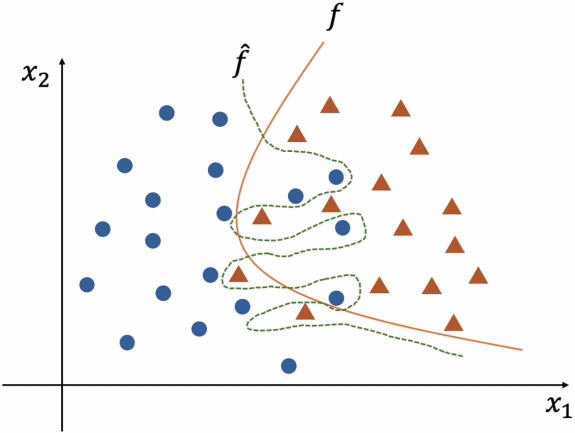

- The usual tradeoff between bias and variance.
- The drop in population risk with decrease in empirical risk is counterbalanced by the increase in model (hypothesis class) complexity.

# Underparameterized Regime



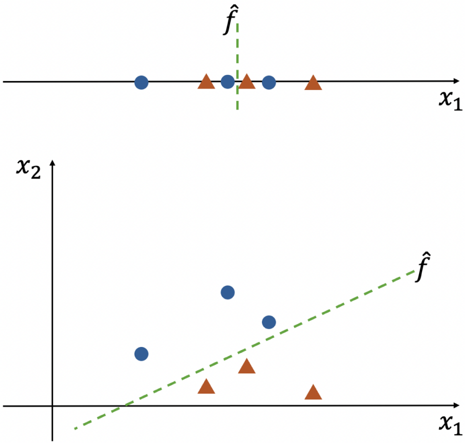

- High bias due to low complexity of the hypothesis class.

# Underparameterized Regime




- High variance due to high complexity of the hypothesis class.

# Overparameterized Regime Intuition




- Adding more dimensions gives the classifier higher flexibility to separate data. Hypothesis class becomes richer.
- Binary classification problem becomes linearly separable in higher dimensions.

# Overparameterized regime

- However, in overparameterized regime, classifiers often have no unique solution to minimizing empirical loss.
- The loss function has many global minima i.e. loss equals 0. Most of them do not generalize well.
- Optimization algorithm matters! Certain algorithms have implicit bias that lead to a global minimum that generalizes well. E.g. gradient descent with squared loss in overparameterized linear regression leads to the ridgeless interpolator.
- Soudry et al (2018) show that in usual classification, linear classifier with gradient descent converges to the maximum hyperplane for linearly separable data.

# Convergence to SVM

## Proposition (Convergence to SVM with Weighted Loss)

*For any dataset that is linearly separable, suppose surrogate loss function $\phi$ and learning rate $\eta$ satisfy the conditions in Theorem 3 of Soudry et al (2018). For a linear classifier $f(\mathbf{x_i}) = \mathbf{x_i}^T w$, let $w(t)$ be the weight at iteration t of gradient descent, for both $\mathcal{L}(w) = \frac{1}{m}\sum_{i=1}^{m} \phi(t_i sign(\mathbf{x_i}^T w))$ and $\mathcal{L}^{\omega}(w) = \frac{1}{m}\sum_{i=1}^{m} \omega_i \phi(t_i sign(\mathbf{x_i}^T w))$, the following holds:*

$$\lim_{t\to\infty} \frac{w(t)}{||w(t)||} = \frac{\hat{w}}{||\hat{w}||} \tag{10}$$

*for almost all datasets (all except measure zero) and strictly positive $\omega_i$. $\hat{w}$ is the direction of the max-margin SVM*

# Convergence to SVM Intuition

- Gradient flow:

$$-\nabla\mathcal{L}(w(t)) = \sum_{i=1}^{m} \omega_i \exp(-g(t) w_\infty^T \mathbf{x_i} y_i) \cdot \exp(-\rho(t)^T \mathbf{x}_n y_i)\mathbf{x_n} y_i$$

- When data is separable, $g(t) \to \infty$, $\frac{\rho(t)}{g(t)} \to 0$, so only those samples with the largest exponents e.g. smallest $g(t) w_\infty^T \mathbf{x_i} y_i$ will contribute to the gradient, i.e. the support vectors.
- Max margin SVM solves $\min ||w||^2$ s.t. $w^T \mathbf{x_i} y_i \geq 1 \ \forall i$. Weighted loss is like replicating the constraints, which does not change the support vectors.

# VC Dimension Bound for SVM

- Vapnik (2000) Theorem 5.1, let $d$ be the VC dimension of a max margin hyperplane:

$$d \leq \min\{\frac{R^2}{\rho^2}, n\} + 1$$

where $\mathbf{x_i}$ is assumed bounded in a sphere of radius $R$, $\rho$ is the margin of the SVM, and $n$ is the dimension of the space.

- For max margin SVM:

$$\rho^2 = \frac{1}{||\hat{w}||^2}$$

# Decreasing Complexity in Overparameterized Regime

In the overparamerized regime, empirical welfare is maxed out, so population welfare can be written as:

$$W(f) \approx C - O\left(\sqrt{\log m}\left(\sqrt{\frac{1 + \log \frac{m}{||\hat{\mathbf{w}}||^2}}{\frac{m}{||\hat{\mathbf{w}}||^2}}}\right)\right)$$

# Decreasing Complexity

- For max margin SVM, $||\hat{\mathbf{w}}||$ is non-increasing in the dimension of $\mathbf{x_i}$, otherwise the additional dimension can have coefficient 0 in the minimization problem.
- In the domain where $||\hat{\mathbf{w}}|| \leq \sqrt{m}$, $\sqrt{\frac{1+\log\frac{m}{||\hat{\mathbf{w}}||^2}}{\frac{m}{||\hat{\mathbf{w}}||^2}}}$ is strictly decreasing in $\frac{m}{||\hat{\mathbf{w}}||^2}$.
- First ascent due to bias variance trade-off, second ascent entirely due to complexity control without explicit regularization.
- Relevant explanations: Belkin et al (2020), Deng et al (2020), Kini and Thrampoulidis (2020), Hastie et al (2022).

# Overview

# Takeaway for Practitioners

- In certain situations, e.g. you believe the true DGP is high dimensional or a low dimensional set of latent features gives many noisy measurements, practitioners may benefit from overparameterized models because richer feature sets can lead to higher welfare level than more parsimonious specifications.
- Besides linear model with gradient descent, can also consider neural network classifiers. Although shown to work well in practice, theoretical reasons for generalization of NN is an open question.

# Some Complications

- Interpretability.
- Equity.
- Preference for simple decision rules.
- Resource constraints in collecting features.

# Unanswered Questions

- What if using observational data? How does estimating propensity score affect welfare performance?
- How does this compare to existing policy learning methods e.g. Empirical Welfare Maximization, Athey & Wager (2021).
- How does this compare to estimating CATE then plug in CATE for treatment assignment? e.g. deep neural network results of Farrell et al (2021).

Thank you.