EECE-6822 Machine Learning

Bias-Variance Tradeoff
Regularization

Cris Ababei
Dept. of Electrical and Computer Engineering
MARQUETTE UNIVERSITY
BE THE DIFFERENCE.

PART 1

Bias-Variance Tradeoff

Regularization, Ridge Regression

Outline

- Bias-Variance Tradeoff
- Regularization

THE BIAS/VARIANCE TRADE-OFF

An important theoretical result of statistics and machine learning is the fact that a model's generalization error can be expressed as the sum of three very different errors:

Bias

This part of the generalization error is due to wrong assumptions, such as assuming that the data is linear when it is actually quadratic. A high-bias model is most likely to underfit the training data. [6]

Variance

This part is due to the model's excessive sensitivity to small variations in the training data. A model with many degrees of freedom (such as a high-degree polynomial model) is likely to have high variance and thus overfit the training data.

Irreducible error

This part is due to the noisiness of the data itself. The only way to reduce this part of the error is to clean up the data (e.g., fix the data sources, such as broken sensors, or detect and remove outliers).

Increasing a model's complexity will typically increase its variance and reduce its bias. Conversely, reducing a model's complexity increases its bias and reduces its variance. This is why it is called a trade-off.

Optimal Prediction

Goal: Predict $Y \in \mathbb{R}$ **given** $X \in \mathbb{R}^d$ **if** $(X, Y) \sim P_{XY}$

Find function η that minimizes

$$\mathbb{E}_{XY}[(Y - \eta(X))^2] = \mathbb{E}_X\left[\mathbb{E}_{Y|X}[(Y - \eta(x))^2 | X = x]\right]$$

(Hint: for any x, $\eta(x) = c_x$ where c_x minimizes $\mathbb{E}_{Y|X}[(Y - c_x)^2 | X = x]$)

$$\begin{aligned} 0 &= \frac{d}{dc_x}\mathbb{E}_{Y|X}[(Y - c_x)^2 | X = x] \\ &= \mathbb{E}_{Y|X}[\frac{d}{dc_x}(Y - c_x)^2 | X = x] \\ &= \mathbb{E}_{Y|X}[-2(Y - c_x) | X = x] = -2\mathbb{E}_{Y|X}[Y | X = x] + 2c_x \end{aligned}$$

Squared Error Optimal Predictor: $\eta(x) = \mathbb{E}_{Y|X}[Y | X = x]$

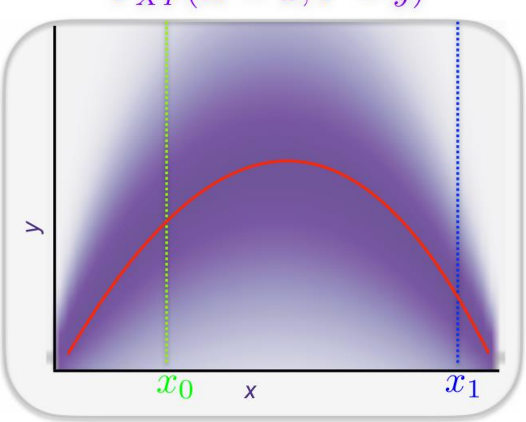
5

5

Statistical Learning

$\mathbb{E}_{XY}[(Y - \eta(X))^2]$

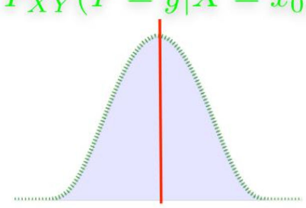
$P_{XY}(X = x, Y = y)$

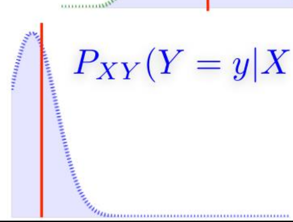


Ideally, we want to find:

$\eta(x) = \mathbb{E}_{Y|X}[Y | X = x]$

$P_{XY}(Y = y | X = x_0)$





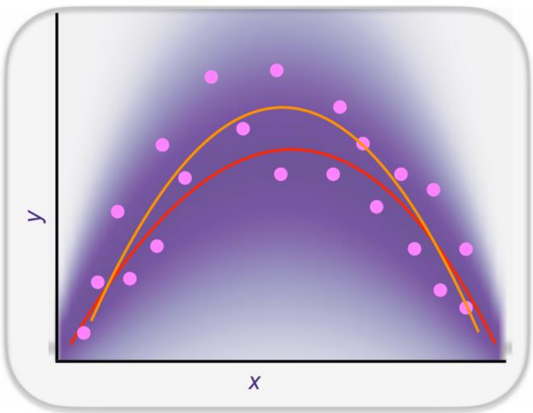
$P_{XY}(Y = y | X = x_1)$

6

6

Statistical Learning

$P_{XY}(X = x, Y = y)$



Ideally, we want to find:

$\eta(x) = \mathbb{E}_{Y|X}[Y|X = x]$

But we only have samples:

$(x_i, y_i) \overset{i.i.d.}{\sim} P_{XY}$ for $i = 1, \ldots, n$

and are restricted to a function class (e.g., linear) so we compute:

$$\widehat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2$$

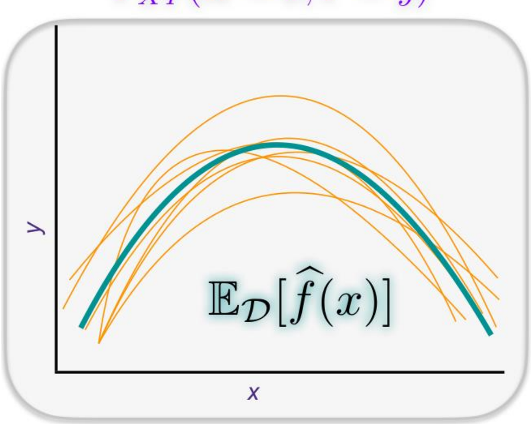
We care about future predictions: $\mathbb{E}_{XY}[(Y - \widehat{f}(X))^2]$

7

7

Statistical Learning

$P_{XY}(X = x, Y = y)$



Ideally, we want to find:

$\eta(x) = \mathbb{E}_{Y|X}[Y|X = x]$

But we only have samples:

$(x_i, y_i) \overset{i.i.d.}{\sim} P_{XY}$ for $i = 1, \ldots, n$

and are restricted to a function class (e.g., linear) so we compute:

$$\widehat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Each draw $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ results in different $\widehat{f}$

8

8

Bias-Variance Tradeoff

$$\eta(x) = \mathbb{E}_{Y|X}[Y|X=x] \qquad \widehat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

$$\mathbb{E}_{Y|X}[\mathbb{E}_{\mathcal{D}}[(Y-\widehat{f}_{\mathcal{D}}(x))^2]\big|X=x] = \mathbb{E}_{Y|X}[\mathbb{E}_{\mathcal{D}}[(Y-\eta(x)+\eta(x)-\widehat{f}_{\mathcal{D}}(x))^2]\big|X=x]$$

$$= \mathbb{E}_{Y|X}\Big[\mathbb{E}_{\mathcal{D}}[(Y-\eta(x))^2 + 2(Y-\eta(x))(\eta(x)-\widehat{f}_{\mathcal{D}}(x)) + (\eta(x)-\widehat{f}_{\mathcal{D}}(x))^2]\big|X=x\Big]$$

$$= \mathbb{E}_{Y|X}[(Y-\eta(x))^2\big|X=x] + \mathbb{E}_{\mathcal{D}}[(\eta(x)-\widehat{f}_{\mathcal{D}}(x))^2]$$

irreducible error
Caused by stochastic label noise

learning error
Caused by either using too "simple" of a model or not enough data to learn the model accurately

9

Bias-Variance Tradeoff

$$\eta(x) = \mathbb{E}_{Y|X}[Y|X=x] \qquad \widehat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

$$\mathbb{E}_{\mathcal{D}}[(\eta(x)-\widehat{f}_{\mathcal{D}}(x))^2] = \mathbb{E}_{\mathcal{D}}[(\eta(x)-\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)]+\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)]-\widehat{f}_{\mathcal{D}}(x))^2]$$

$$= \mathbb{E}_{\mathcal{D}}[(\eta(x)-\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)])^2 + 2(\eta(x)-\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)])(\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)]-\widehat{f}_{\mathcal{D}}(x)) + (\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)]-\widehat{f}_{\mathcal{D}}(x))^2]$$

$$= (\eta(x)-\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)])^2 + \mathbb{E}_{\mathcal{D}}[(\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)]-\widehat{f}_{\mathcal{D}}(x))^2]$$

biased squared **variance**

For more insights read also: Abu-Mostafa 2.2-2.3

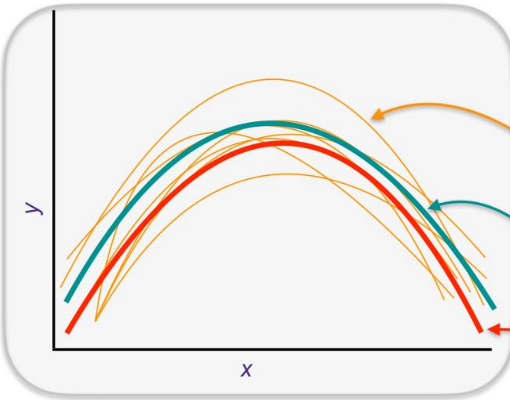
10

10

Statistical Learning

$$\underbrace{\mathbb{E}_{\mathcal{D}}[(\eta(x) - \widehat{f}_{\mathcal{D}}(x))^2]}_{\text{learning error}} = \underbrace{(\eta(x) - \mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)])^2}_{\text{biased squared}} + \underbrace{\mathbb{E}_{\mathcal{D}}[(\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)] - \widehat{f}_{\mathcal{D}}(x))^2]}_{\text{variance}}$$

learning error = biased squared + variance



$$\widehat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2$$

$$\mathbb{E}_{\mathcal{D}}[\widehat{f}(x)]$$

$$\eta(x) = \mathbb{E}_{Y|X}[Y|X = x]$$

11

11

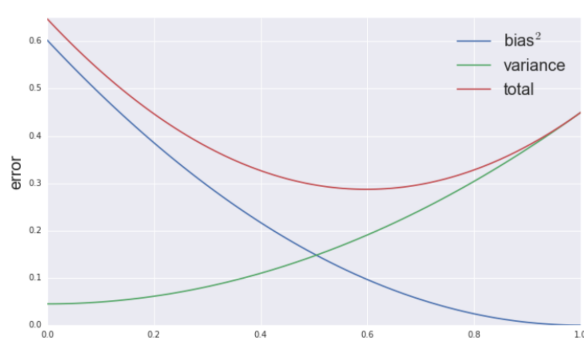
Bias-Variance Tradeoff

$$\mathbb{E}_{Y|X}[\mathbb{E}_{\mathcal{D}}[(Y - \widehat{f}_{\mathcal{D}}(x))^2] \big| X = x] = \mathbb{E}_{Y|X}[(Y - \eta(x))^2 \big| X = x]$$

irreducible error

$$+(\eta(x) - \mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)])^2 + \mathbb{E}_{\mathcal{D}}[(\mathbb{E}_{\mathcal{D}}[\widehat{f}_{\mathcal{D}}(x)] - \widehat{f}_{\mathcal{D}}(x))^2]$$

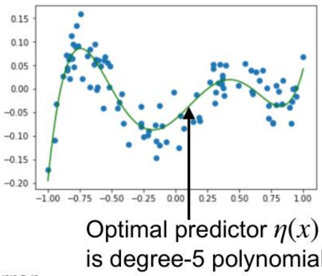
biased squared   variance

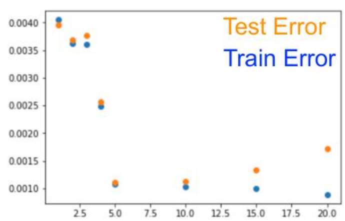


12

12

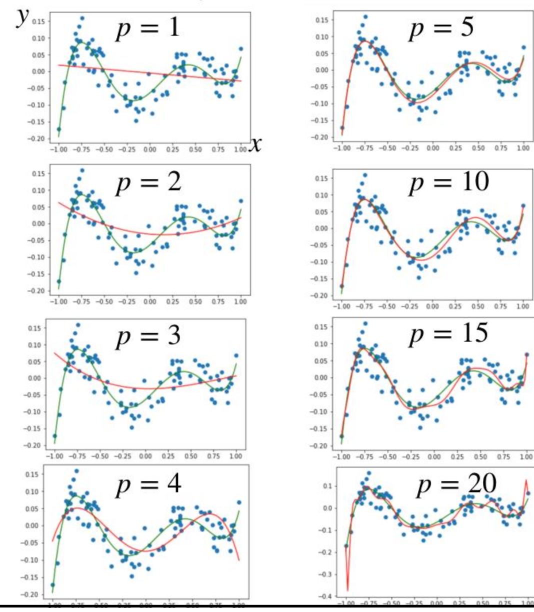
Test error vs. model complexity





Simple model:
Model complexity is below the complexity of $\eta(x)$

Complex model:
Fits noise in train data, diverging from $\eta(x)$

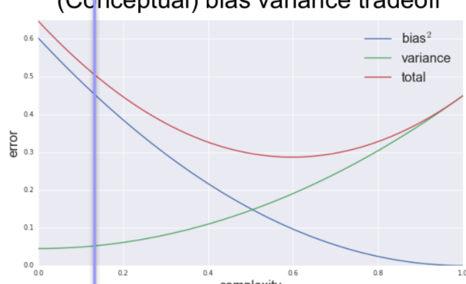


13

13

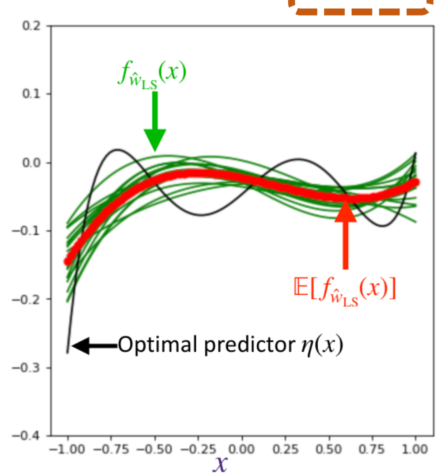
Recap: Bias-variance tradeoff with simple model

(Conceptual) bias variance tradeoff



- When model **complexity is low** (lower than the optimal predictor $\eta(x)$)
  - Bias^2 of our predictor, $\left(\eta(x) - \mathbb{E}_{\mathcal{D}}[\hat{f}_{\mathcal{D}}(x)]\right)^2$, is large
  - Variance of our predictor, $\mathbb{E}_{\mathcal{D}}\left[\left(\mathbb{E}_{\mathcal{D}}[\hat{f}_{\mathcal{D}}(x)] - \hat{f}_{\mathcal{D}}(x)\right)^2\right]$, is small
  - **If we have more samples (larger n), then**
    - What happens to bias?
    - What happens to variance?
    - What happens to overall test error?

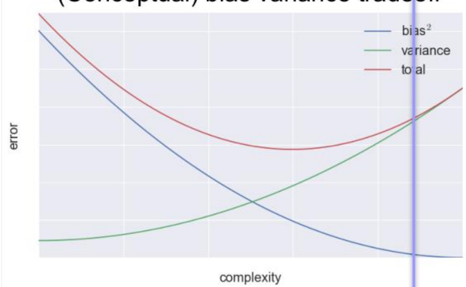
With degree-3 polynomials, we underfit

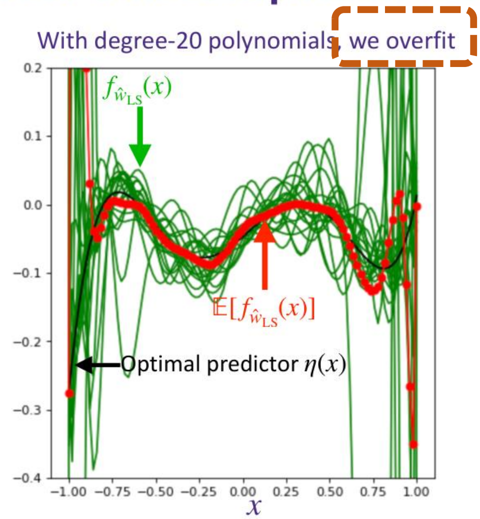


14

14

Recap: Bias-variance tradeoff with complex model





- When model complexity is high (higher than the optimal predictor $\eta(x)$)
  - Bias of our predictor, $\left(\eta(x) - \mathbb{E}_{\mathcal{D}}[\hat{f}_{\mathcal{D}}(x)]\right)^2$, is small
  - Variance of our predictor, $\mathbb{E}_{\mathcal{D}}\left[\left(\mathbb{E}_{\mathcal{D}}[\hat{f}_{\mathcal{D}}(x)] - \hat{f}_{\mathcal{D}}(x)\right)^2\right]$, is large
- **If we have more samples (larger n), then**
  - What happens to bias?
  - What happens to variance?
  - What happens to overall test error?

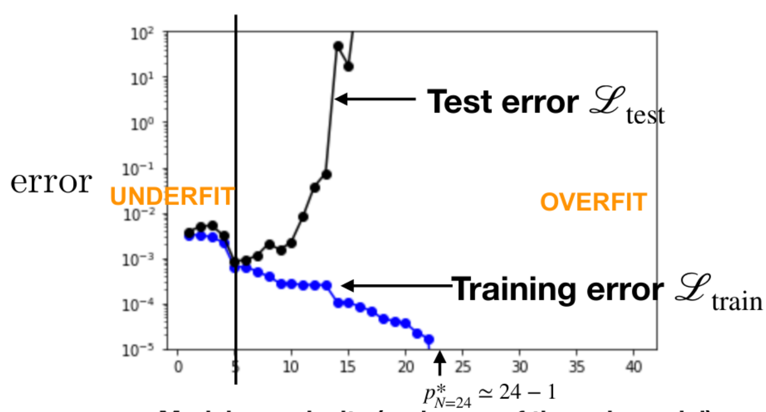
15

15

Optimal model complexity depends on dataset size

Takeaway

- Assume ***N=30***



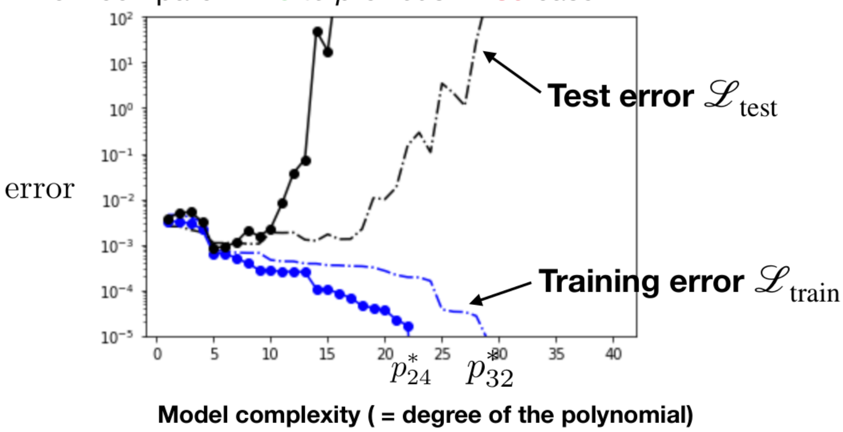
- Given sample size N there is a threshold, p^*_N, where training error is zero
- Training error is **always** monotonically non-increasing
- Test error has a trend of going down and then up, but fluctuates

16

16

Variance decreases with more data, letting you fit more complex models

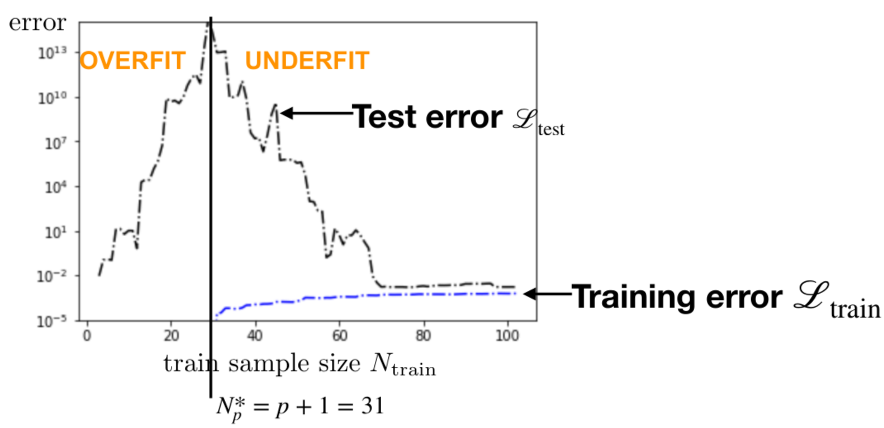
- Now compare ***N=40*** *to previous* ***N=30*** *case*



- **The threshold, p^*_N, moves right as dataset size increases**
- Training error tends to increase, because more points need to fit
- Test error tends to decrease, because Variance decreases

Variance decreases with more data, letting you fit more complex models

- Choose model complexity ***p=30***, **vary dataset size n**



- There is a threshold, N^*_p, below which training error is zero (extreme overfit)
- Above the threshold, test error tends to decrease
- Training error tends to increase (harder to fit so much data with simple model)

Regularization - Helps Avoid Overfitting
Ridge Regression

19

19

Regularization in Linear Regression

Recall Least Squares: $\widehat{w}_{LS} = \arg\min_w \sum_{i=1}^{n} \left(y_i - x_i^T w\right)^2$

$$= \arg\min_w (\mathbf{y} - \mathbf{X}w)^T (\mathbf{y} - \mathbf{X}w)$$

when $(\mathbf{X}^T\mathbf{X})^{-1}$ exists.... $= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

What if $x_i \in \mathbb{R}^d$ and $d > n$?

20

20

Regularization in Linear Regression

Recall Least Squares: $\widehat{w}_{LS} = \arg\min_{w} \sum_{i=1}^{n} \left(y_i - x_i^T w\right)^2$

When $x_i \in \mathbb{R}^d$ and $d > n$ the objective function is flat in some directions:

Implies optimal solution is *not unique* and unstable due to lack of curvature:

- small changes in training data result in large changes in solution
- often the *magnitudes* of *w* are "very large"

Regularization imposes "simpler" solutions by a "complexity" penalty

21

Sensitivity increases overfitting

- For a linear model,
  $y \simeq b + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d$
  if $|w_j|$ is large then the prediction is **sensitive** to small changes in x_j
- Large ***sensitivity*** leads to overfitting and poor generalization, and equivalently models that overfit tend to have large weights
- Note that b is a constant and hence there is no sensitivity for the offset b
- In ***Ridge Regression***, we use a regularizer $\|w\|_2^2$ to measure and control the sensitivity of the predictor
- And optimize for small loss and small sensitivity, by adding a ***regularizer*** in the objective (assume no offset for now)

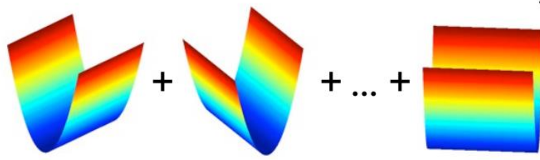
$$\hat{w}_{\text{ridge}} = \arg\min_{w} \left\{ \sum_{i=1}^{n} (y_i - x_i^T w)^2 + \lambda \|w\|_2^2 \right\}$$

2

Ridge Regression

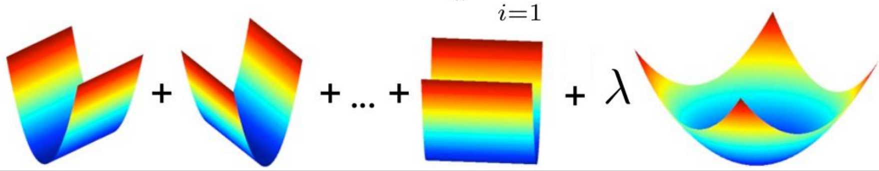
- Old Least squares objective:

$$\widehat{w}_{LS} = \arg\min_w \sum_{i=1}^n \left(y_i - x_i^T w\right)^2$$



- Ridge Regression objective:

$$\widehat{w}_{ridge} = \arg\min_w \sum_{i=1}^n \left(y_i - x_i^T w\right)^2 + \lambda ||w||_2^2$$



Minimizing the Ridge Regression Objective

$$\widehat{w}_{ridge} = \arg\min_w \sum_{i=1}^n \left(y_i - x_i^T w\right)^2 + \lambda ||w||_2^2$$

Scalar derivative			Vector derivative		
$f(x)$	$\rightarrow$	$\frac{df}{dx}$	$f(\mathbf{x})$	$\rightarrow$	$\frac{df}{d\mathbf{x}}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{B}$	$\rightarrow$	$\mathbf{B}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{b}$	$\rightarrow$	$\mathbf{b}$
x^2	$\rightarrow$	$2x$	$\mathbf{x}^T\mathbf{x}$	$\rightarrow$	$2\mathbf{x}$
bx^2	$\rightarrow$	$2bx$	$\mathbf{x}^T\mathbf{B}\mathbf{x}$	$\rightarrow$	$2\mathbf{B}\mathbf{x}$

Same idea as before: take derivative, set to zero. Arrive at:

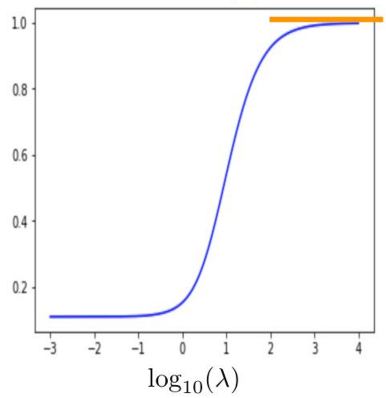
$$\widehat{w}_{ridge} = \arg\min_w \sum_{i=1} \left(y_i - x_i^T w\right)^2 + \lambda ||w||_2^2$$

$$= (\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^T\mathbf{y}$$

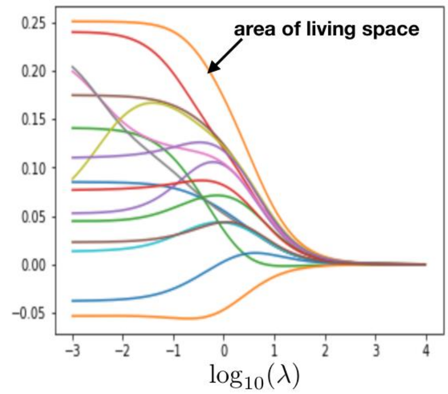
- When $\lambda = 0$, this gives the least squares model
- This defines a family of models hyper-parametrized by λ
- Large λ means more regularization and simpler model
- Small λ means less regularization and more complex model

Ridge regression: minimize $\sum_{i=1}^{n} (w^T x_i - y_i)^2 + \lambda \|w\|_2^2$

training MSE $\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T \hat{w}_{\text{ridge}}^{(\lambda)})^2$



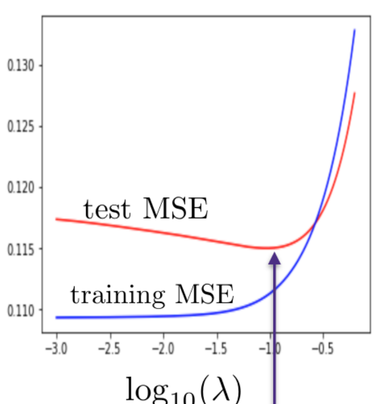
Housing price predictor $w_i's$



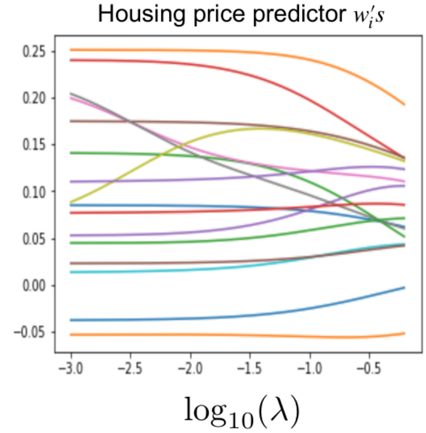
- Left plot: leftmost training error is with no regularization: 0.1093
- Left plot: rightmost training error is variance of the training data: 0.9991
- Right plot: called **regularization path**

25

Ridge regression: minimize $\sum_{i=1}^{n} (w^T x_i - y_i)^2 + \lambda \|w\|_2^2$



Housing price predictor $w_i's$



- this gain in test MSE comes from shrinking w's to get a less sensitive predictor
(which in turn reduces the variance)

- this is the role of regularizer

26

Role of Regularizer

- **Penalizes large coefficients:**
  - It discourages the model from relying too heavily on any single feature by shrinking weights toward zero.
- **Controls model complexity:**
  - By constraining weight size, it reduces the flexibility of the regression function, which helps prevent overfitting.
  - Higher complexity = huge coefficients that wiggle to fit noise; more flexible model → more complex decision boundary
  - Lower complexity = smaller closer-to-each-other coefficients means simpler, smoother models

27

27

Bias-Variance Properties

- Recall: $\hat{w}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$
- To analyze bias-variance tradeoff, we need to assume probabilistic generative model: $x_i \sim P_X, \;\; \mathbf{y} = \mathbf{X}w + \epsilon, \;\; \epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$
- The true error at a sample with feature x is

$$\mathbb{E}_{y,\mathcal{D}_{\text{train}}|x}[(y - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$
$$= \mathbb{E}_{y|x}[(y - \mathbb{E}[y\,|\,x])^2 \,|\, x] + \mathbb{E}_{\mathcal{D}_{\text{train}}}[(\mathbb{E}[y\,|\,x] - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$
$$= \mathbb{E}_{y|x}[(y - x^Tw)^2 \,|\, x] + \mathbb{E}_{\mathcal{D}_{\text{train}}}[(x^Tw - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$
$$= \underbrace{\sigma^2}_{\text{Irreduc. Error}} + \underbrace{(x^Tw - \mathbb{E}_{\mathcal{D}_{\text{train}}}[x^T\hat{w}_{\text{ridge}} \,|\, x])^2}_{\text{Bias-squared}} + \underbrace{\mathbb{E}_{\mathcal{D}_{\text{train}}}[(\mathbb{E}_{\tilde{\mathcal{D}}_{\text{train}}}[x^T\hat{w}_{\text{ridge}} \,|\, x] - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]}_{\text{Variance}}$$

Suppose $\mathbf{X}^T\mathbf{X} = n\mathbf{I}$, then $\hat{w}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T(\mathbf{X}w + \epsilon)$

$$= \frac{n}{n+\lambda}w + \frac{1}{n+\lambda}\mathbf{X}^T\epsilon$$

28

28

Bias-Variance Properties

Suppose $\mathbf{X}^T\mathbf{X} = n\mathbf{I}$, then

$$\hat{w}_{\text{ridge}} = \frac{n}{n+\lambda} w + \frac{1}{n+\lambda}\mathbf{X}^T\epsilon$$

- Recall: $\hat{w}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$
- To analyze bias-variance tradeoff, we need to assume probabilistic generative model: $x_i \sim P_X, \ \ \mathbf{y} = \mathbf{X}w + \epsilon, \ \ \epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I})$
- The true error at a sample with feature x is

$$\mathbb{E}_{y,\mathcal{D}_{\text{train}}|x}[(y - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$

$$= \mathbb{E}_{y|x}[(y - \mathbb{E}[y\,|\,x])^2 \,|\, x] + \mathbb{E}_{\mathcal{D}_{\text{train}}}[(\mathbb{E}[y\,|\,x] - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$

$$= \mathbb{E}_{y|x}[(y - x^T w)^2 \,|\, x] + \mathbb{E}_{\mathcal{D}_{\text{train}}}[(x^T w - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$

$$= \sigma^2 + (x^T w - \mathbb{E}_{\mathcal{D}_{\text{train}}}[x^T\hat{w}_{\text{ridge}} \,|\, x])^2 + \mathbb{E}_{\mathcal{D}_{\text{train}}}[(\mathbb{E}_{\tilde{\mathcal{D}}_{\text{train}}}[x^T\hat{w}_{\text{ridge}} \,|\, x] - x^T\hat{w}_{\text{ridge}})^2 \,|\, x]$$

(verify at home)

$$= \sigma^2 \ + \ \frac{\lambda^2}{(n+\lambda)^2}(w^T x)^2 \ + \ \frac{\sigma^2 n}{(n+\lambda)^2}\|x\|_2^2$$

Irreduc. Error — Bias-squared — Variance

9

29

Bias-Variance Properties

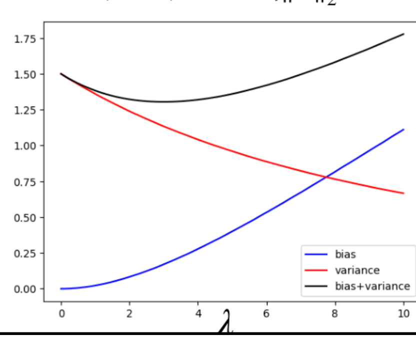
Suppose $\mathbf{X}^T\mathbf{X} = n\mathbf{I}$,

- Ridge regressor: $\widehat{w}_{ridge} = \arg\min_w \sum_{i=1}^n \left(y_i - x_i^T w\right)^2 + \lambda\|w\|_2^2$
- True error

$$\mathbb{E}_{y,\mathcal{D}_{\text{train}}|x}[(y - x^T\hat{w}_{\text{ridge}})^2 \,|\, x] = \sigma^2 + \frac{\lambda^2}{(n+\lambda)^2}(w^T x)^2 + \frac{\sigma^2 n}{(n+\lambda)^2}\|x\|_2^2$$

Bias-squared — Variance

d=10, n=20, $\sigma^2 = 3.0, \|w\|_2^2 = 10$



as λ →0, $\hat{w}_{\text{ridge}} \to \hat{w}_{\text{LS}}$

as λ →∞ $\hat{w}_{\text{ridge}} \to 0$

30

30

Recap

- Regularization
  - Penalizes complex models towards preferred, simpler models
- Ridge regression
  - L_2 penalized least-squares regression
  - Regularization parameter trades off model complexity with training error
  - Never regularize the offset!

31

31

Sparsity & the LASSO

How to make the model compact and interpretable

See lecture notes

32

32

PART 2

More Details

Details

- See hand-written notes and discussion in class.

PART 3

Code Time!

35

35

Code Time

- See demonstration and discussion in class.
- See also links in the “Assignment” for this Lecture.
- Bias Variance Tradeoff
  - bias_variance_demo.ipynb (uwash)
- Regularization, Ridge Regression
  - ridge_lasso_kaggle.ipynb (uwash)

36

36

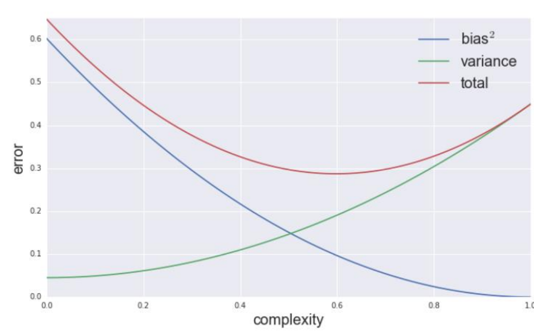
Conclusion
Takeaways

37

37

Conclusion

- **Learning error** = Biased Squared + Variance
  - Caused by either using too “simple” of a model or not enough Data to learn the model accurately
- **Regularization**
  - Imposes “simpler” solutions by a “complexity” penalty



38

38

References and Credits

Many of the teaching materials for this course have been adapted from various sources. We are very grateful and thank the following professors, researchers, and practitioners for sharing their teaching materials (in no particular order):

- Yaser S. Abu-Mostafa, Malik Magdon-Ismail and Hsuan-Tien Lin. https://amlbook.com/slides.html
- Ethem Alpaydin. https://www.cmpe.boun.edu.tr/~ethem/i2ml3e/
- Natasha Jaques. https://courses.cs.washington.edu/courses/cse446/25sp/
- Lyle Ungar. https://alliance.seas.upenn.edu/~cis520/dynamic/2022/wiki/index.php?n=Lectures.Lectures
- Aurelien Geron. https://github.com/ageron/handson-ml3
- Sebastian Raschka. https://github.com/rasbt/machine-learning-book
- Trevor Hastie. https://www.statlearning.com/resources-python
- Andrew Ng. https://www.youtube.com/playlist?list=PLoROMvodv4rMiGQp3WXShtMGgzqpfVfbU
- Richard Povineli. https://www.richard.povinelli.org/teaching
- … and many others.

39