CS 146 Discussion Week 2

Linear Basis Function Models, Generalization

Outline

- Linear Basis Function Models
- Generalization
  - Regularization
  - Model Selection
- We will use this notebook during the discussion:
  - https://colab.research.google.com/drive/1Lur43Fz9lkK8oNvf97GYSQaEsaKPl1rK?usp=sharing
  - Notebook is adapted from https://jakevdp.github.io/PythonDataScienceHandbook/05.06-linear-regression.html

Linear Basis Function Models

- Generalized class of linear hypothesis

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}^T \phi(\boldsymbol{x}) = \sum_{j=0}^{k} \theta_j \phi_j(\boldsymbol{x})$$

- $\phi(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}^k$ is a k-dimensional basis with parameters $\boldsymbol{\theta} \in \mathbb{R}^k$

- Polynomial basis functions (polynomial regression) with degree k

$$\phi_j(x) = x^j$$

$$h_\theta(x) = \sum_{j=0}^{k} \theta_j \phi_j(x) = \theta_0 + \theta_1 x + \ldots + \theta_k x^k$$

- Gaussian basis functions (we use k of them)

$$\phi_j(x) = e^{-\frac{1}{2}(\frac{x-\mu_j}{\sigma})^2}, \phi_0(x) = 1$$

$$h_\theta(x) = \sum_{j=0}^{k} \theta_j \phi_j(x)$$

Regularized Linear Regression

- Loss Function: $$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{d}\theta_j^2$$
- Gradient Descent Update (note no regularization on bias term)

$$\theta_0 \leftarrow \theta_0 - \alpha\frac{1}{n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)$$

$$\theta_j \leftarrow \theta_j - \alpha\frac{1}{n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)x_j^{(i)} - \alpha\lambda\theta_j$$

- Closed Form Solution (exercise if you have time: derive out that the inverse always exists for positive $\boldsymbol{\lambda}$)

$$\boldsymbol{\theta} = \left(\boldsymbol{X}^\intercal\boldsymbol{X} + \lambda\begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}\right)^{-1} \boldsymbol{X}^\intercal\boldsymbol{y}$$

Linear Regression Exercise

- Given training dataset below, compute the closed-form linear regression solution.

i	1	2	3	4	5
$x^{(i)}$	-2	-1	0	1	2
$y^{(i)}$	1	3	5	7	9

- Formula for closed form solution:

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Linear Regression Exercise

- Put all data points into vectorized form and prepend 1 (for the bias term)

$$\boldsymbol{X} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}$$

- Apply formula $\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$
- $\boldsymbol{\theta} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

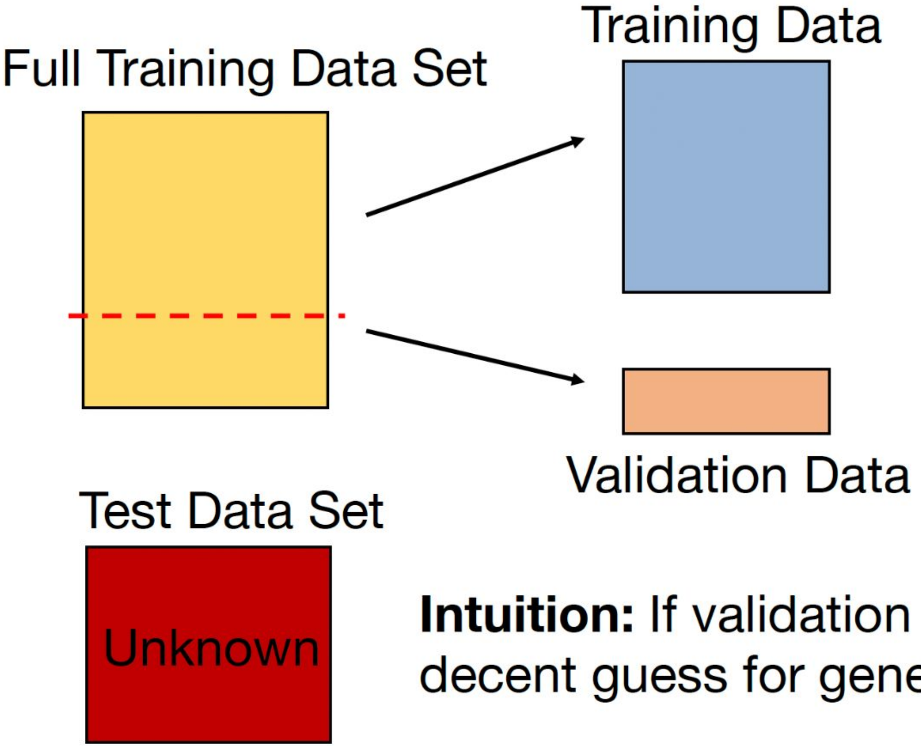
Regularized Linear Regression Exercise

- Same dataset, compute closed-form solution for regularized linear regression (ridge regression) with $\lambda = 1$
- Apply formula

$$\boldsymbol{\theta} = \left(\boldsymbol{X}^\intercal \boldsymbol{X} + \lambda \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \right)^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}$$

- $$\boldsymbol{\theta} = \begin{bmatrix} 5 \\ 1.82 \end{bmatrix}$$
- The bias term (5) stays the same
- Exercise (if you have time): show that bias term does not change from linear regression to ridge regression. This is true because we do not regularize the bias term.

Training, Validation, Test data



Idea:
Train each model on a subset of full training data...

...and then evaluate each model's accuracy on the **validation data** a.k.a. held-out data

Intuition: If validation data ~ test data, then we have a decent guess for generalization

Cross-Validation