

## 1. Summary:

Over the first two weeks of the quarter, we covered an overall summary of what statistical learning is and broke it down into supervised and unsupervised learning. We went over statistical machine learning's goals of prediction/inference and using  $\hat{f}$  to predict  $y$  in the context of linear regression. We also discussed MSE and the tradeoff between bias (from underfitting to training data) and variance (from overfitting), and proving the unbiasedness of  $\beta_1/\beta_0$  in linear regression. We understood how to obtain the parameters in linear regression and how they are calculated.

## 2. Concepts:

Statistical Machine Learning:

- Supervised Machine Learning - predicting a specific outcome based on features/predictors
- Unsupervised Machine Learning - Understanding the relationship between features, more exploratory
- Input can be a vector with values for each feature
- Model is:  $y = f(X) + \epsilon$
- $x$ 's are observations of predictors and  $y$ 's are observations of outputs
- Goal: for a given  $x$ , predict  $y$ ,  $\hat{f}(x)$  is the estimate corresponding to a given  $x$
- Prediction: predict  $\hat{y}$  as close to true value of  $y$  as possible
- Inference: Understand the underlying mechanism, relationship in the data (ex. which features best predict  $y$ )
- $E[Y|X] = \operatorname{argmin}_g E[(y - g(x))^2 | X=x]$
- $E[(Y - f(X))^2 | X=x] = [f(x) - \hat{f}(x)]^2 + \operatorname{Var}(\epsilon)$
- Left part is reducible while right is not as it is random from  $\epsilon$
- $\text{MSE for a specific } X=x \geq \operatorname{Var}(\epsilon)$
- There are trade offs in model flexibility and interpretability
- To assess model accuracy use new test set (data we have not seen before in training)
- $\text{MSE} = E[\operatorname{var}(\hat{f}(x)) + \operatorname{Bias}(\hat{f}(x))^2] + \operatorname{Var}(\epsilon)$
- High variance is a result of overfitting to training data (more complex model)
- High bias is a result of underfitting to training data (more simple model)
- Achieving an ideal trade off is very important

Linear Regression:

- Absolute test MSE is useless, relative MSE is important and can be compared

- Process:
  - Compute  $\hat{f}$  for every model class  $C$
  - Compare different  $\hat{f}$  based on test MSE and choose  $\hat{f}$  with smallest test MSE
  - Simple LR:  $Y = \beta_0 + \beta_1 X + \varepsilon$
- Goal is to effectively estimate  $\hat{\beta}_0$  and  $\hat{\beta}_1$
- Residual sum of squares is amount of variability not explained by our model
  - Also mean squared error
- To find  $\hat{\beta}$  you want to minimize RSS (the MSE)
- You want to differentiate it find  $\hat{\beta}$  that satisfies zero gradient
  - Check notes for proof and more detail on specific formulas
- Assessing model fit
  - $R^2 = (TSS - RSS)/TSS$
  - $TSS \geq RSS$
  - $R^2$  between 0 and 1
  - $R^2 = (\text{Correlation}(X, Y))^2$
- Prove  $\hat{\beta}_1$  and  $\hat{\beta}_0$  are unbiased estimators of their respective parameters
- $SE_{\hat{\beta}_1} = \sqrt{(RSS/n-1)/\sum(x_i - \bar{x})^2)}$
- $(\hat{\beta}_1 - \beta_1)/SE$  is a  $t$  dist  $n - 2$  df
- Confidence Interval for  $\beta_1$ :  $[\hat{\beta}_1 - t SE_{\hat{\beta}_1}, \hat{\beta}_1 + t SE_{\hat{\beta}_1}]$
- If 0 not in CI reject null hypothesis
- Type 1 error =  $1 - \text{confidence percent} = 1 - 0.95 = 0.05$

### 3. Uncertainties:

Some areas I need to delve deeper into include looking at the proofs deeper for proving why  $\hat{\beta}_1$  is an unbiased estimator - specifically seeing why the gradient we used is necessary as well as the matrix operations we used. Since they were such a small portion of the work we did, it went over my head a little bit, and I need to understand it further. Overall, the major ideas all make complete sense to me, but there are some intermediate steps I need to look over again to ensure I understand how we reached the results. I also want to understand the matrix operations in multiple regression (which we will cover next week). Also, the single value decomposition we went over in section did not make too much sense to me because I had forgotten the exact meaning of eigenvectors, so this is something I need to study and look through thoroughly as well. Understanding more of the weeds and details of what we did will help me grasp the overall concepts even better in the coming weeks of the class. I am looking forward to moving on to new concepts and getting started with classification.