

## CMSE 381 - Quiz #1

①  $f(x) = E(Y|X=x) = \arg \min_{g(x)} E[(Y-g(x))^2 | X=x]$

② Bayes Classifier:  $C(x) = j$  if  $P_j(x) = \max \{P_1(x), P_2(x), \dots, P_K(x)\}$



③ This is unsupervised learning because there is no outcome variable, just a set of predictors.

It it were supervised learning, there would be a clear outcome variable, & that is not the case here.

④ (a) The Red curve is the testing MSE. This because the testing MSE, unlike the training MSE, can never go to 0. It will at most drop to the dashed line, which is  $\text{Var}(\epsilon)$ , the irreducible error.

(b) The dashed line at 1.0 represents the  $\text{Var}(\epsilon)$ , which is the irreducible error in the model.

⑤ (a) Variance ( $\text{Var}(\hat{y}_0)$ ) is basically a measurement of how much the model would fluctuate if it were estimated on a different training set.

Bias is basically, on average, after fitting models with multiple training datasets, how much the average prediction differs from the true value/data.

(b) The tuning parameter  $\lambda$  is basically a way to make models less flexible & more rigid. This is typically done to prevent models from overfitting the data and following noise instead of signal. This restriction decreases variance but because of the bias-variance tradeoff, the bias will increase as the model predicts further away from the true value.



⑥  $f(\text{Salary}) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Residence}_{\text{Michigan}} + \beta_3 \text{Residence}_{\text{Ohio}} \leftarrow \text{above}$

$$\text{Salary} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Residence}_{\text{Michigan}} + \beta_3 \text{Residence}_{\text{Ohio}} + \epsilon$$

10

if  $y = \text{Salary}$ ,  $x_1 = \text{Age}$ ,  $x_2 = \text{Residence}_{\text{Michigan}}$ ,  $x_3 = \text{Residence}_{\text{Ohio}}$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon$$

$\beta_1$  is an average, how much a year increase in age will affect the salary of an individual, holding their residence constant

$\beta_2$  is an average how much more a Michigan resident makes <sup>in salary</sup> relative to the Indiana baseline, holding age constant

$\beta_3$  is an average how much more an Ohio resident makes in salary to the Indiana baseline, holding age constant.

$\beta_0$  is the baseline, return to how much an Indiana resident makes, holding age constant.