

CMSE QUIZ 1

① continuous RV

40

$$\int y \cdot f(y) \rightarrow f_y \cdot f_{x,y}(x, y)$$

5 $L(Yf(x)) = (Y - f(x))^2$

oracle function

$$E(Y | X=x)$$

② using the indicator function

Bayes classifier

$$j_i \cdot \Pr(Y \neq C(X_i))$$

{own}

2

$$\begin{cases} 1 & Y \neq C(X_i) \\ 0 & Y = C(X_i) \end{cases}$$

③ This is unsupervised learning because we don't have the labels

In unsupervised learning, we don't have the Y . Here we only have observed characteristics [nothing more to work from].

④

- a) The orange curve is testing for the MSE. The other curve is the curve looking at bias

5

The orange curve appears to be MSE_{TE} because it never crosses the dashed line

- b) The dashed line is the optimal model [you would never be able to achieve because of irreducible error]

6

Fill in the following information about the model:
Labels, features, error term, model

⑥

Labels = age, residence, state

4

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

β_1 = age

β_2 = residence (MD, OH, IN)

β_3 = state

⑤ (a) $\text{var}(\hat{f}(x_0))$: Increases when model is flexible

[$\text{Bias}(\hat{f}(x_0))$]² : Decreases when the model is very flexible (overfitting)

(b) If you decrease the value of λ , the bias would increase because with a lower λ , the model becomes less flexible which would increase bias

for