CSPB 4622 - Truong - Machine Learning

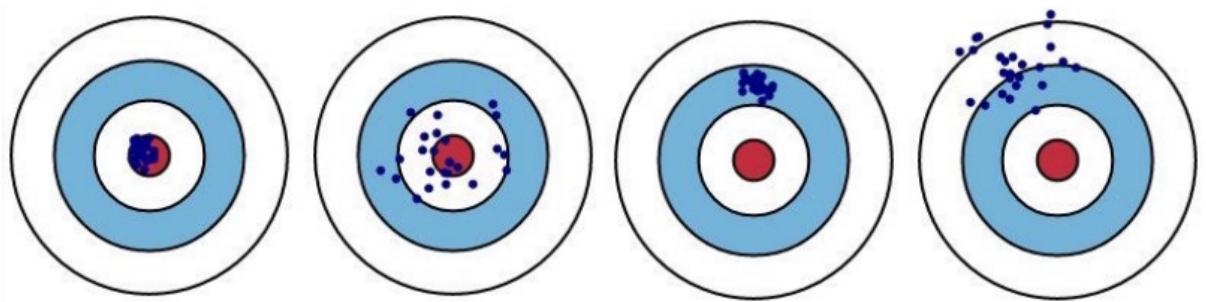|	
Started on	Tuesday, 3 September 2024, 3:33 PM
State	Finished
Completed on	Tuesday, 3 September 2024, 3:40 PM
Time taken	7 mins 43 secs
Marks	7.00/7.00
Grade	**10.00** out of 10.00 (**100**%)

Question **1**

Correct

Mark 1.00 out of 1.00

How would you characterize the third model from the left?

- ○ a. High bias and high variance
- ○ b. Low bias and high variance
- ◉ c. High bias and low variance ✔
- ○ d. Low bias and low variance

Question **2**

Correct

Mark 1.00 out of 1.00

Using a subset of the Iris dataset, a regression model of **PetalLength ~ PetalWidth * C(kind)** was run, where **kind** is a categorical variable taking on values 'versicolor' or 'virginica'. The resulting model has an $R^2 = 0.74$ and parameters as shown in the following table:

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.7813	0.413	4.309	0.000	0.961	2.602
C(kind)[T.virginica]	2.4594	0.614	4.005	0.000	1.241	3.678
PetalWidth	1.8693	0.308	6.061	0.000	1.257	2.482
PetalWidth:C(kind)[T.virginica]	-1.2221	0.380	-3.215	0.002	-1.976	-0.468

Calculate the predicted **PetalLength** given a Virginica iris with PetalWidth of 2. You only need to worry about two digits of precision (*e.g.* treat the intercept as 1.78).

Answer: 4.92 ✔

Question **3**

Correct

Mark 1.00 out of 1.00

Assume you have data where the number of predictors **p** is extremely large and the number of observations **n** is small. Would you generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method?

- ○ a. Better than inflexible method.
- ◉ b. Worse than an inflexible method.   ✔

Question **4**

Correct

Mark 1.00 out of 1.00

The below result is a multilinear regression fit on the data with variables shown below. The crime rate is the dependent variable (Y), and the rest are predictors (X).

Which variables/terms can you remove from the model? We use the standard significance threshold of 0.05.

Choose all that apply.

Variable	Description
crim	per capita crime rate by town
zn	proportion of residential land zoned for lots over 25,000 sq.ft.
indus	proportion of non-retail business acres per town.
rm	average number of rooms per dwelling.
dis	weighted mean of distances to five Boston employment centers.
rad	index of accessibility to radial highways.
black	1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.
lstat	lower status of the population (percent).
mdev	median value of owner-occupied homes in $1000s.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	5.2286	4.748	1.101	0.271	-4.100	14.557
zn	0.0469	0.017	2.697	0.007	0.013	0.081
indus	-0.1487	0.070	-2.112	0.035	-0.287	-0.010
rm	0.4081	0.601	0.679	0.498	-0.773	1.589
dis	-0.7778	0.242	-3.210	0.001	-1.254	-0.302
rad	0.4879	0.044	11.011	0.000	0.401	0.575
black	-0.0078	0.004	-2.128	0.034	-0.015	-0.001
lstat	0.1289	0.071	1.808	0.071	-0.011	0.269
medv	-0.1610	0.056	-2.893	0.004	-0.270	-0.052

☐

a. Intercept

☐

b. zn

☐

c. indus

☑

d. rm ✔

☐

e. dis

☐

f. rad

☐

g. black

☑

h. lstat ✔

☐

i. medv

Question **5**

Correct

Mark 1.00 out of 1.00

When a true relationship between X and Y is linear, which of the below is True?

○ a. RSS of a simple model $Y = aX + b$ on the training dataset is about the same as the RSS of a cubic model $Y = aX + bX^2 + cX^3 + d$ on the training dataset.

○ b. RSS of a simple model $Y = aX + b$ on the test dataset is larger than the RSS of a cubic model $Y = aX + bX^2 + cX^3 + d$ on the test dataset.

○ c. There is not enough information.

◉ d. RSS of a simple model $Y = aX + b$ on the training dataset is larger than the RSS of a cubic model $Y = aX + bX^2 + cX^3 + d$ on ✔ the training dataset.

Question **6**

Correct

Mark 1.00 out of 1.00

Below table shows results from two multilinear regression models. What are the reasonable conclusions?

|		Coefficient	Standard Error	p-value
Model 1	Intercept X1 X2	-150.6 -2.52 0.352	39.6 0.55 0.007	<0.0001 0.0003 <0.0001
Model 2	Intercept X2 X3	-340.9 0.037 2.74	42.5 0.082 1.02	<0.0001 0.72 0.022

☐

d(a) There is nonlinearity among features

☑

b(b) There is a strong correlation between the feature X2 and X3

✔

☑

d(c) There is collinearity among features

✔

☐

d(d) Feature X2 can be removed from both models because its coefficient is negligible and unreliable

Question **7**

Correct

Mark 1.00 out of 1.00

Which is the most appropriate term for curve D?
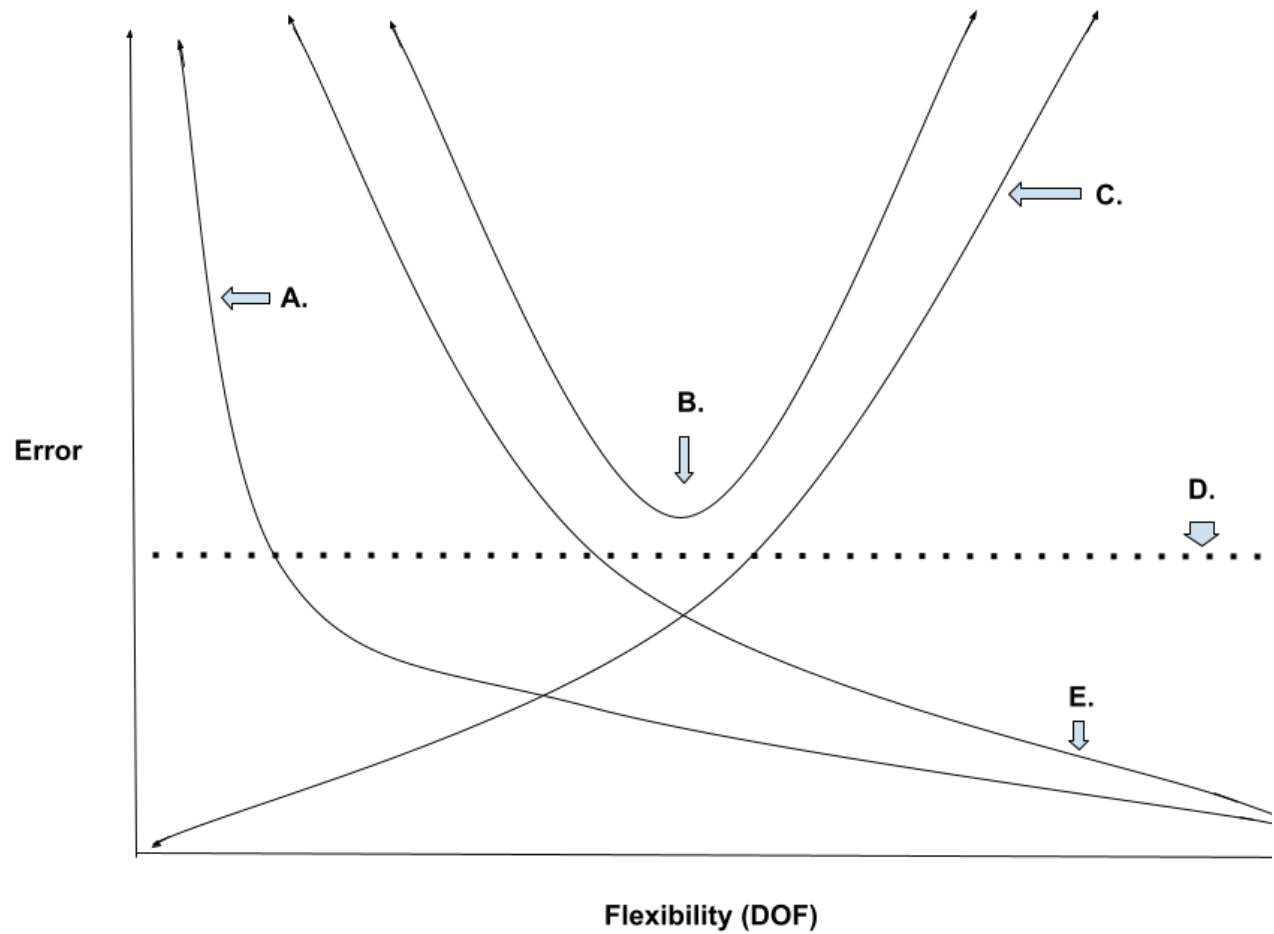
a. Bias

b. Variance

c. Training error

d. Test error

e. Bayes' or irreducible error ✔