Quiz 3.1

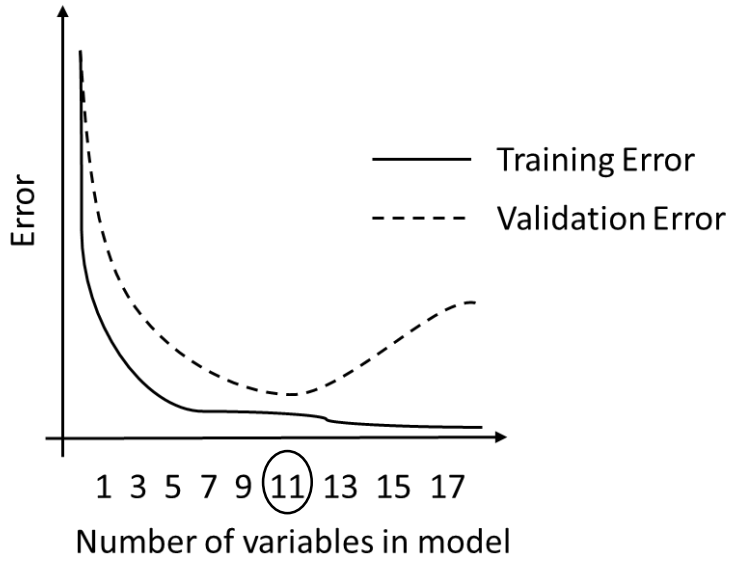
1.) **(2 pts.)** Bias and variance are two different types of error for a predictive model. Circle the term on the left that matches the definition on the right.

Bias / **Variance**	Error due to a model's inability to replicate the fundamental phenomena represented by a data set, or the error defined as: $\boldsymbol{E}\left[\hat{\boldsymbol{f}}(\boldsymbol{x})\right] - \boldsymbol{f}(\boldsymbol{x})$**.** where $\boldsymbol{E}$**[]** is the expected value, or average, operator**,** $\boldsymbol{f}(\boldsymbol{x})$, is the true value of the target variable for a data set, and $\hat{\boldsymbol{f}}(\boldsymbol{x})$ is the predicted value of the target variable for a data set.
Bias / **Variance**	Error due to a model's ability to produce differing predictions from the values in a new data set, or the error defined as $\boldsymbol{E}[\hat{\boldsymbol{f}}(\boldsymbol{x}) - \boldsymbol{E}[\hat{\boldsymbol{f}}(\boldsymbol{x})]]^{\boldsymbol{2}}$

2.) **(3 pts.)** Per common nomenclature, circle the term on the left that matches the definition on the right.

Training / validation / test data	A partition of data used for fitting the model parameters or building the rules that define the model.
Training / validation / test data	A partition of data used only for the final, honest estimate of model performance.
Training / **validation** / **test** data	A partition of data used for model selection and tuning the model to improve its generalization.

3.) **(1 pt.)** In the figure below, two partitions of data are used to fit and asses a predictive model. The variance of the model is increased, and the bias of the model is decreased, by adding predictor variables into the model. At what number of variables does the model display the best generalization abilities given only the information in the figure? 11



A linear regression analysis was conducted to determine the relationship between the amounts a hospital charges for a medical service (AVE_ave_provider_charge), the amount a hospital is reimbursed by Medicare (AVE_ave_medicare_payment), and the number of services a hospital provides (AVE_num_service).

The model formula was specified as:

AVE_ave_provider_charge ~ AVE_ave_medicare_payment + AVE_num_service

Among many other tables and plots, the following information was provided by the statistical software package after training the traditional regression model:

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1127.54	563.92	-2.00	0.0416
AVE_ave_medicare_payment	Average Medicare Payment	1	4.03	0.09	45.44	<.0001
AVE_num_service	Number of Services	1	-4.11	1.19	-3.45	<.0001

R-Square	**0.657**
Adj R-Sq	0.649

4.) **(2 pts.)** State the exact interpretation of the presented standard R-Square statistic.

The trained linear model explains 65.7% of the variance in the response variable, AVE_ave_provider_charge.

5.) **(2 pts.)** State the exact interpretation of the presented parameter estimate for AVE_num_service.

Holding all other variables constant, a 1-unit increase in the average number of services (AVE_ num_service) will result in the average amount charged by a provider (AVE_ave_provider_charge) decreasing by 4.11.