# DATA 100: Vitamin 9 Solutions

April 9, 2019

## 1 Risk Optimization

Let $\hat{\theta}_n(x_0)$ be an estimator for $\theta(x_0) = E[Y|X = x_0]$ using a randomly sampled learning set $X_n$. The mean squared error (MSE) of this estimator is equal to:

- ☑ $E[(\hat{\theta}_n(x_0) - \theta(x_0))^2|X_n]$
- ☐ $Var(\hat{\theta}_n(x_0)|X_n) + E[\hat{\theta}_n(x_0)|X_n]$
- ☐ $E[(E[\hat{\theta}_n(x_0)|X_n] - \theta(x_0))^2]$
- ☑ $Var(\hat{\theta}_n(x_0)|X_n) + Bias(\hat{\theta}_n(x_0)|X_n)^2$

**Explanation:** See the proof in lecture 17's slides. Always remember to condition on $\boldsymbol{X}_n$!

## 2 Bias-Variance Trade-Off

### 2.1 Underfitting

Fill in the blanks:

Underfitting occurs when a model is too ____ to represent the _____ of the learning set.

- ☑ simple, data generating mechanism
- ☐ complex, data generating mechanism
- ☐ simple, bias
- ☐ complex, bias

### 2.2 Underfitting

Overfitting occurs when a model is too ____ and fits the ____ from sampling in the learning set.

- ☐ simple, variance
- ☐ complex, variance
- ☐ simple, noise
- ☑ complex, noise

**Explanation (Q2.1 and Q2.2):** Underfitting consists of fitting a model that is too simple, and overfitting consists of fitting a model that is too complex. When a model is overly complex, it tends to fit to the noise in the data instead of the signal (the true data generating mechanism). Striking a balance between simplicity and complexity is not an easy task.

### 2.3 Complexity

More complex models can reduce bias, but can increase variance. Which of the following are example ways to measure model complexity?

- ☑ the number of parameters
- ☑ the length of the parameter vector
- ☑ the number of covariates in a linear regression function
- ☑ the degree of the polynomial in a polynomial regression function
- ☐ None of these

**Explanation:** See the lecture slides for a list of models and their corresponding measures of complexity.

## 3 Cross Validation

K-fold cross-validation (CV) consists of partitioning the ____ set data into K disjoint subsets of roughly equal size, which are then used to train estimators and evaluate their performance. If using CV for model selection, the model with the ____ CV risk is selected. Using CV for model selection will help you avoid selecting a model that overfits to the noise in the learning data.

- ☐ test, highest
- ☐ learning, highest

☐ test, lowest

☑ learning, lowest

**Explanation:** CV consists of partitioning the learning set. When using CV for model selection, the model with the lowest CV risk is chosen.