DATA 100: Vitamin 3 Solutions

September 13, 2019

1 Data Science Life Cycle

The data science life cycle broadly consists of four elements. In no particular order, they are:

- ☑ Question formulation, data analysis, data design/generation, and generalization
- ☐ Data analysis, data procedures, learning, and visualization
- ☐ Machine learning, artificial intelligence, data design/generation, hypothesis testing

Explanation: The data science life cycle consists of question formulation, data design and generation, data analysis, and generalization. Data design and generation, and data analysis are the elements that we will be focusing on in this course.

2 Loss Functions

The ℓ_2 loss function for some data point x_i and an estimate c is defined as:

- ☐ $|x_i - c|$
- ☐ $(x_i - c)$
- ☐ $\sqrt{(x_i - c)}$
- ☑ $(x_i - c)^2$

Explanation: $\ell_2(x_i, c) = (x_i - c)^2$

3 Sampling

Fill in the blank:
A ____ is a random sample where every member of the sampling frame has the same chance of being in the sample.

- ☑ simple random sample
- ☐ stratified sample
- ☐ cluster sample

Explanation: The statement defines a simple random sample.

4 0-1 Random Variables

Which of the following events can be modeled using 0-1 random variables? Select all that apply.

- ☑ Rolling a fair dice and obtaining a 6.
- ☑ Randomly selecting a King from a deck of cards.
- ☐ The number of hurricanes to form in the Atlantic ocean in any given year.
- ☑ Determining how many of your family members voted for Trump in the 2016 election.

Explanation: The first two options are 0-1 random variables since each event only has two random outcomes: success or failure. The third event is not a 0-1 random variable since there are more than two outcomes. Instead, we might consider using a Poisson distribution to model the number of hurricanes. The final option can be modeled using a collection of 0-1 random variables. We call this a binomial distribution.

5 Variance of Random Variables

Fill in the blank:
$Var(aX + b) =$ ____ , where a, b are scalars and X is a random variable.

- ☐ $aVar(X) + b$
- ☐ $a^2Var(X) + b$
- ☐ $aVar(X)$
- ☑ $a^2Var(X)$

Explanation:

$$\begin{aligned} Var(aX + b) &= E[(aX + b - E(aX + b))^2] \\ &= E[(aX + b - (aE(X) + b))^2] \\ &= E[(aX - aE(X))^2] \\ &= E[a^2(X - E(X))^2] \\ &= a^2E[(X - E(X))^2] \\ &= a^2Var(X) \end{aligned} \tag{1}$$