P2

P2 :

Proof: Number that minimizes mean squared error is the mean of the dataset.

$$f(c) = \frac{1}{N}\left[(x_1 - c)^2 + (x_2 - c)^2 + \ . \ . \ . \ + (x_N - c)^2\right]$$

$$f'(c) = \left[\frac{d}{dc}\left[(x_1 - c)^2\right] + \frac{d}{dc}\left[(x_2 - c)^2\right] \ \cdots \right)\frac{1}{N}$$

$$\downarrow$$

$$\frac{d}{dc}\left[x_1^2 - 2x_1c + c^2\right]$$

$$= \frac{1}{N}\left(\left[2c - 2x_1\right] + \left[2c - 2x_2\right] + \ . \ . \ . \left[2c - 2x_n\right]\right)$$

$$f'(c) = \frac{1}{N}\left(2Nc - 2\left[x_1 + x_2 + \ . \ . \ . + x_N\right]\right)$$

$$f'(c) = 2c - \frac{2}{N}[x_1 + x_2 + \ldots + x_N]$$

$$f'(c) = 2\left[c - \frac{1}{N}[x_1 + x_2 + \ldots + x_N]\right]$$

$$f'(c) = 0 = 2\left[c - \frac{1}{N}[x_1 + x_2 + \ldots + x_N]\right]$$

$$0 = c - \frac{1}{N}[x_1 + x_2 + \ldots + x_N]$$

$$\boxed{c = \frac{1}{N}[x_1 + x_2 + \ldots + x_N]}$$

$$c = \frac{1}{4}\left(175 + 172 + 180 + 185\right)$$

$\boxed{c = 178}$

P3

$Y = f(X) + \varepsilon$, ε is independent noise

Minimize

$$E\left[(Y - \hat{f}(x))^2\right]$$

$$E\left[\left(\log(x) + \varepsilon - \hat{f}(x)\right)^2\right]$$

$$\hat{Y} = \log X + c$$

$$E\left[\left(\log(x) + \varepsilon - (\log(x) + c)\right)\right]$$

$$E\left[(\varepsilon - c)^2\right]$$

What value of c best describes random noise of ε?

As proved in P2, the number that minimizes the squared error loss between c and ε is the mean of ε, $\bar{\varepsilon}$.

Therefore, the function, $\hat{f}(X)$, that best predicts the output is

$$\hat{f}(x) = \log(x) + \bar{\varepsilon}$$

or

$$\boxed{\hat{f}(x) = \log(x) + 10}$$

P4

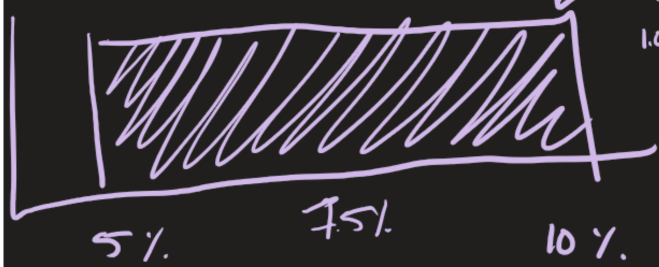
With more features comes a higher dimensionality in the data. This presents issues in the K-Nearest Neighbors algorithm, because the algorithm is reliant on the Euclidean distance formula for calculating the nearest neighbors to a given point. Because the formula adds another term for each dimension of the vector, when distance is calculated for high dimensional inputs, the nearest neighbors to a given data point may not be close at all.

P5

Stratified sampling is not needed for sufficiently large enough datasets. This is because the Law of Large Numbers will allow for uniform sampling to represent features equally. However, this does **not** apply to small datasets due to the high variance that will appear in small datasets.

P6

$$\int_{1.05}^{1.1} f(x)\,dx = 1$$



$$E(X) = \int_{1.05}^{1.1} x f(x)\,dx$$

$$f(x) = \frac{1}{1.1 - 1.05} = \frac{1}{0.05} = 20$$

$$= \int_{1.05}^{1.1} x \cdot 20\,dx$$

$$= 20 \int_{1.05}^{1.1} x\,dx$$

$$= 20 \left. \frac{x^2}{2} \right|_{1.05}^{1.1}$$

$$= 10 \left[x^2 \Big|_{1.05}^{1.1} \right]$$

$$= 10 \left[1.1^2 - 1.05^2 \right]$$

$$= 10 \left[0.1075 \right]$$

$$= 1.075$$

Expected inflation is the average, or 7.5%.

Because there is an equal chance of staying 1 or 2 years, we can find the expected value by using the probabilities and expected costs like so:

$$\left(\tfrac{1}{2}\right)(1.075)(10{,}000) + \left(\tfrac{1}{2}\right)\left[(1.075)(10{,}000) + (1.075)^2(10{,}000)\right] = \text{Expected Total Cost}$$

$$5{,}375 + 11{,}153.125 = \text{Expected Total Cost}$$

$16,528.125 = Expected Total Cost