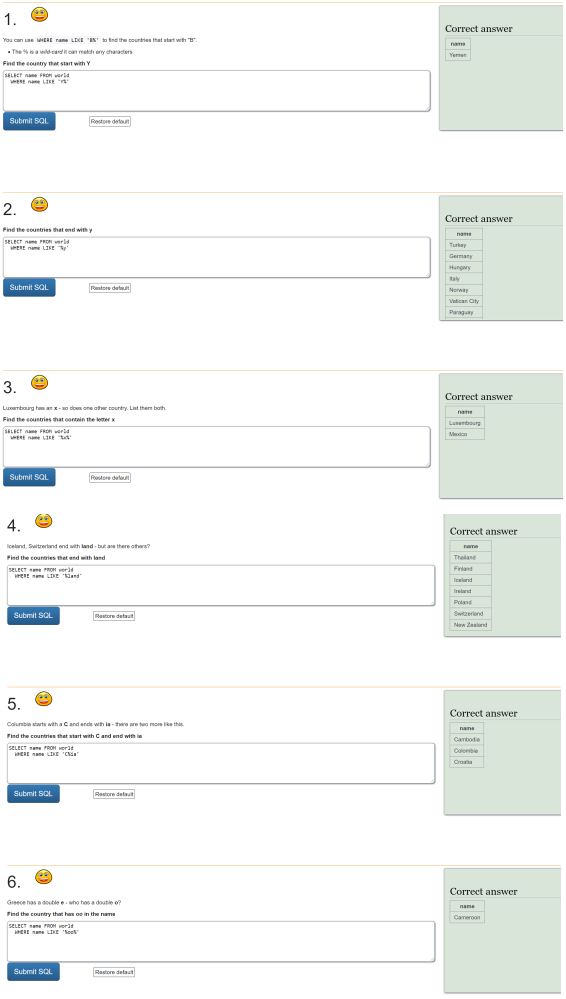
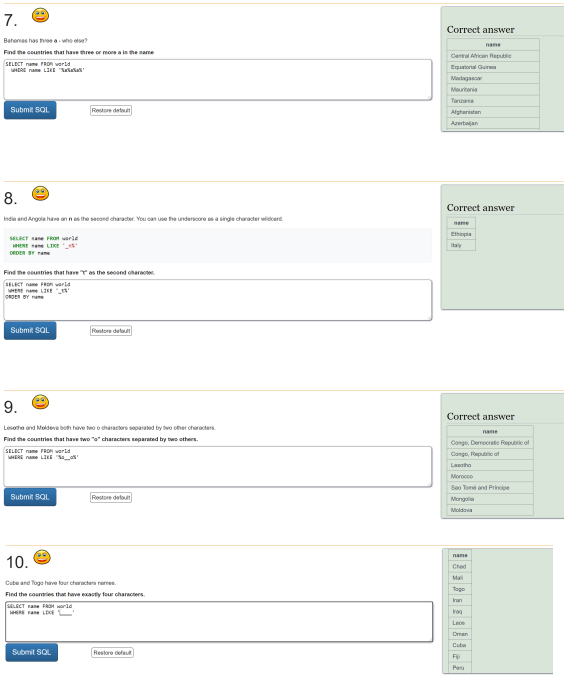
Machine Learning Assignment 64

Elijah Tarr

February 24, 2021

Problem 2





Problem 3

(a)

$$\text{PMF of Y} = \begin{cases} 0.2 & \text{for } k \in [0,1,2] \\ 0 & \text{otherwise} \end{cases}$$

(b)

$$\begin{aligned} R_X &= [0,2] \\ P(X \geq 1.5) &= \frac{1}{6} \\ P(0 < X < 2) &= \frac{1}{3} \\ P(X = 0 | X < 2) &= \frac{P(X = 0 \cap X < 2)}{P(X < 2)} \\ &= \frac{3}{5} \end{aligned}$$

(c)

$$
\begin{aligned}
P(A-C) &= \frac{11}{12} - \frac{3}{4} \\
&= \frac{1}{6} \\
P(B-C) &= \frac{11}{12} - \frac{2}{3} \\
&= \frac{1}{4} \\
P(C) &= \frac{11}{12} - \frac{1}{6} - \frac{1}{4} \\
&= \frac{1}{2} \\
P(A \cap C) &= P(A)P(C) \\
x &= \frac{(x + \frac{1}{6})}{2} \\
P(A \cap C) &= \frac{1}{6} \\
P(A) &= \frac{1}{3} \\
P(B \cap C) &= P(B)P(C) \\
y &= \frac{(y + \frac{1}{4})}{2} \\
P(B \cap C) &= \frac{1}{4} \\
P(B) &= \frac{1}{2}
\end{aligned}
$$

(d)

$$
\begin{aligned}
P(A) &= P(B) \\
P(C) &= 2P(D) \\
x &= P(A) \\
y &= P(C) \\
x + y &= 0.6 \\
x + \frac{y}{2} &= 1 - 0.6 \\
&\text{Solve system of equations:} \\
x &= 0.2 \\
y &= 0.4 \\
P(A) &= 0.2 \\
P(B) &= 0.2 \\
P(C) &= 0.4 \\
P(D) &= 0.2
\end{aligned}
$$

Problem 4

(a)

Due to overfitting, we would expect the cubic regressor to have a smaller RSS value than the linear regressor. This is because the cubic regressor would have extra terms and coefficients that allow it to account slightly more for the random epsilon values added to each data value. The linear regressor would just fit the best straight line to the data, which will likely be further away from some points than the cubic regressor. However, seeing as the true relationship between the points is linear, we would expect the RSS from a different set of data to be larger for the cubic regressor than the linear regressor. This is again, because the cubic regressor overfits to the certain dataset it's given, but won't exactly follow any dataset as well. So technically, the answer is the former since we are told that the RSS's compared are from the training dataset.

(b)

As explained in answer (a), the linear regressor's RSS is expected to be lower.

(c)

For this one, there isn't enough information to tell. We can be sure that it is a curve, since our regressors will only add terms with higher orders. For example, if the highest order term is x^2, the linear regressor will have a lower RSS. Imagine the true relationship is modeled $y = x^2$. We would have lines $y = x$ and $y = x^3$ for our linear and cubic regressors. Notice how where $x < 0$, the cubic line decreases drastically faster than the straight line, and also increases much faster than the straight line where $x > 0$. However, if the true relationship has an odd order, the cubic regressor will be better, since it will simulate the line on both sides of the y axis. All this is assuming that the data

we are given can be in any quadrant, but if we are only given data where $x > 0$, it is more likely that the cubic regressor works better than linear, since the other side of the y axis won't matter.

(d)

Again, there wouldn't be enough information, as explained in (c).