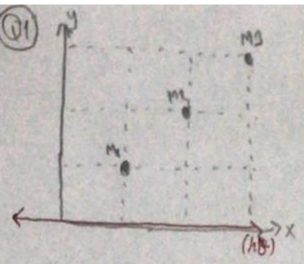
Mert Tazeoğlu
21946606

(Q1)



$H(\theta) = \theta_1 x$

$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$

$m=3, \theta_1=0 \ldots$

$h(\theta) = 0 \cdot x$ (x 0x(i))

$J(0) = \frac{1}{2m}\sum_{i=1}^{m=3} h_\theta(x^{(i)}) - y^{(i)})^2$

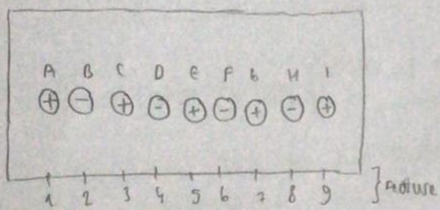
For $m=1 \rightarrow (h_\theta(1) - 0)^2 = (1-0)^2 = 1$

For $m=2 \rightarrow (h_\theta(2) - 0)^2 = (2-0)^2 = 4$

For $m=3 \rightarrow (h_\theta(3) - 0)^2 = (3-0)^2 = 9$

$J(0) = \frac{1}{2m}(1+4+9)^2 = \frac{1}{6}\cdot(1+4+9) = \frac{14}{6} = 7/3$

(Q2)



(10 dataset, in other words, its like a line)

SAMPLE	ACTUAL	PREDICTED
A	+	−
B	−	+
C	+	−
D	−	+
E	+	−
F	−	+
G	+	−
H	−	+
I	+	−

Since currently selected samples nearest neighbour's value (class) is always returned in the 1NN classification algorithm, if the class of each sample's neighbours is the opposite of / different from the current sample's class, the error value for 1-NN with LOOCV will be 1, in other words the classification algorithm won't be able to guess correct class for any specific point in the dataset.

(Q3) X_2 = square of MT exam

$\text{mean}(X_2) = \frac{7569 + 4900 + 8464 + 4489 + 2025}{5} = 5489.9$

$\text{range}(X_2) = 8464 - 2025 = 6439$

$X_2^{(2)} = 4900$

$\text{normalized } X_2^{(2)} = \frac{X_2^{(2)} - \text{mean}(X_2)}{\text{range}(X_2)} = \frac{4900 - 5489.9}{6439} = -0.09$

Q4

$$P(C \mid A, B) = \frac{P(C) \cdot P(A \mid C) \cdot P(B \mid C)}{P(C) \cdot P(A \mid C) \cdot P(B \mid C) + P(C') \cdot P(A \mid C') \cdot P(B \mid C')} \Big\} \text{ required things to calculate}$$

a) $P(C \mid A) = 0.9 = \frac{P(C, A)}{P(A)} = \frac{P(A \mid C) \cdot P(C)}{P(A)}$

$P(C \mid B) = 0.4 = \frac{P(C, B)}{P(B)} = \frac{P(B \mid C) \cdot P(C)}{P(B)}$

we can't compute it with given information. In order to compute it properly, we need to know main prob. values of $P(A)$, $P(B)$ and $P(C)$.

b) $P(A) = 0.3$

$P(B) = 0.5$

We can't compute it with given information. In order to compute it properly we need to know value of $P(C)$.

c) $P(C, A) = 0{,}2$ $\quad P(A) = 0.3$ $\quad P(B) = 1$

$P(C \mid A, B) = P(C \mid A)$ since $P(B) = 1$

Therefore, according to Bayes theorem $P(C \mid A, B) = P(C \mid A) = \frac{P(C, A)}{P(A)} = \frac{0{,}2}{0{,}3} = 2/3$ $\Big\}$ so, we can compute it with given information.

Q5

a) $L(w) = \prod_{i=1}^{m} P_w(x_i)$ where $P_w(x_i) = \begin{cases} 1/2w, & \text{if } x_i \in [-w, w] \\ 0, & \text{otherwise} \end{cases}$

b) If x_i is outside $[-w, w]$ interval then whole likelihood will be 0. At the same time, we would like "w" to be small as possible, in order to maximize $1/(w2)$. Hence, we should pick w to fit both the min and max data point that we are given: $\quad w = \max\left(\left|\min_{1 \le i \le m}(x_i)\right|, \left|\max_{1 \le i \le m}(x_i)\right|\right)$

c) Suppose that you have all the x for the positive class between 0 and some number z and all those for the negative class between $-z$ and 0. Then the machine learning estimate for x for both classes will end up with $w = z$ and as a result, you would always classify the example based on the prior over classes, yet the examples are perfectly separable. The problem here stems from the bias of classifier, which is required to be symmetric.

d) Yes, you could pick a threshold and classify based on whether x is larger or smaller than the threshold. Note that one could also allow non-symmetric boundaries on the distributions, in which case the generative model would work as well.

!

Q6

a) $P(C=1 \mid X=1, Y=1, Z=0) = \dfrac{P(C=1, X=1, Y=1, Z=0)}{P(X=1, Y=1, Z=0)} = \dfrac{P(C=1) \cdot P(X=1 \mid C=1) \cdot P(Y=1 \mid C=1) \cdot P(Z=0 \mid C=1)}{P(X=1, Y=1, Z=0, C=1) + P(X=1, Y=1, Z=0, C=0)}$

$$= \frac{P(C=1) \cdot P(X=1 \mid C=1) \cdot P(Y=1 \mid C=1) \cdot P(Z=0 \mid C=1)}{P(C=1) \cdot P(X=1 \mid C=1) \cdot P(Y=1 \mid C=1) \cdot P(Z=0 \mid C=1) + P(C=0) \cdot P(X=1 \mid C=0) \cdot P(Y=1 \mid C=0) \cdot P(Z=0 \mid C=0)}$$

$$= \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{4} \cdot \frac{1}{8}} = \frac{\frac{1}{32}}{\frac{2}{32}} = \frac{1}{2}$$

b) $P(C=0 \mid X=1, Y=1) = \dfrac{P(C=0, X=1, Y=1)}{P(X=1, Y=1)} = \dfrac{P(C=0) \cdot P(X=1 \mid C=0) \cdot P(Y=1 \mid C=0)}{P(C=0) \cdot P(X=1 \mid C=0) \cdot P(Y=1 \mid C=0) + P(C=1) \cdot P(X=1 \mid C=1) \cdot P(Y=1 \mid C=1)}$

$$= \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{4}} = \frac{\frac{1}{8}}{\frac{1}{8} + \frac{1}{16}} = \frac{2}{3}$$

c) In the given dataset, there isn't any exactly matching records with given input (sample). Therefore:

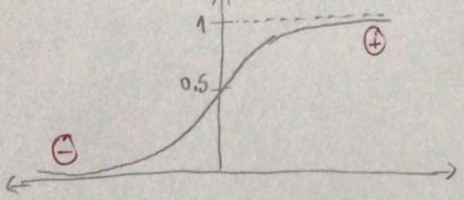
$P(C=1 \mid X=1, Y=1, Z=0) = 0$

d) $P(C=0 \mid X=1, Y=1) = \dfrac{\text{number of } (C=0, X=1, Y=1)}{\text{number of } (X=1, Y=1)} = \dfrac{1}{2}$

Q7

original hypothesis function for logistic regression alg. is:

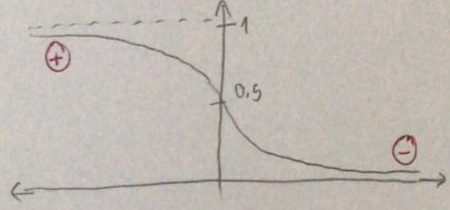
$$h_\theta(x) = \sigma(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$



updated hypothesis function for logistic regression alg. is:

$$h_\theta(x) = \sigma(\theta^T x) = \frac{1}{1+e^{\theta^T x}}$$

(-) sign changed



⇒ Explanation:

New graph has become symmetrical with respect to y-axis. Therefore whatever classification it normally makes now will be opposite. For example if it classifies as (+), now it will classify it as (-). Therefore it can't be used for classification task without making a few changes. For example, if we will use this updated formula, than we should add a line of code about returning opposite version of predicted class in order to make ML model work properly.