Submitted by: Ranji Raj

Assignment 2.1

Solve and discuss the following two subtasks:

(a) Assign the given models (*Logistic Regression*[1], *Perceptron*[2], *ID3 Decision Tree*[3], *K-Means*[4]) to their Machine Learning categorization scheme, e.g. *supervised*, *unsupervised*, *classification*, *regression* and *clustering*.

Supervised	***Unsupervised***	***Classification***	***Regression***	***Clustering***
ID3 Decision Tree	K-Means	ID3 Decision Tree	ID3 Decision Tree	K-Means
Perceptron		Perceptron		
Logistic Regression		Logistic Regression		

(b) Derive a non-formal justification to apply the following *Loss Functions*[5] (*0-1 Loss Function* and the *Sum-Of-Squared-Error Loss*) within the classification and regression setting.

For ***classification***, we have 0-1 loss,
From Bernoulli's theorem for the binary classification problem, we can write,

$$f(y;p) = p^{y} \cdot (1-p)^{1-y}$$

; where y ∈ [0,1] is the ***class label*** & p ∈ [0,1] is the ***predicted probability***

For a binary classifier,

When ***y = 1***, $f(1;p) = p^{1} \cdot (1-p)^{1-1} = p$ &
When ***y = 0,*** $f(0;p) = p^{0} \cdot (1-p)^{1-0} = 1-p$

Now since logarithm is a ***strictly monotonically increasing*** function we take logarithm on both sides,

$$log\left(P\left(y \mid x_i\right)\right) = log\left(p_i^{y} \cdot \left(1-p_i\right)^{1-y}\right)$$

$$log\left(P\left(y \mid x_i\right)\right) = log\left(p_i^{y}\right) + log\left(\left(1-p_i\right)^{1-y}\right)$$

$$log\left(P\left(y \mid x_i\right)\right) = y \cdot log\left(p_i\right) + (1-y) \cdot log\left(1-p_i\right)$$

$$\boxed{0-1\ Loss = -\sum_{i \in \{1,..N\}} y.log\left(p_i\right) + (1-y).log\left(1-p_i\right)}$$

*(-ve sign indicates **minimizing** the loss function thereby **maximizing** the log of the probability)*

For a ***regression*** problem, we make use of the Gaussian distribution,

Given by:
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{\left(x_i-\mu\right)^2}{2\sigma^2}}$$

Where μ = sample mean, σ = sample standard deviation, σ^2 = sample variance

$$L\left(\theta, X_1, \ldots X_N\right) \simeq P\left(X=x_1, \ldots x_N \middle| \theta\right)$$

$$(i.i.d) = \prod_{i \in \{1,..N\}} N\left(X_i=x_i \middle| \mu, \sigma^2\right)$$

$$= \prod_{i \in \{1,..N\}} \frac{1}{\left(\sigma \cdot \sqrt{2\pi}\right)} \cdot e^{-\frac{\left(x_i-\mu\right)^2}{2 \cdot \sigma^2}}$$

$$= \left(\frac{1}{\left(\sigma \cdot \sqrt{2\pi}\right)}\right)^N \cdot e^{\left(\sum_{i \in \{1,..N\}} -\frac{\left(x_i-\mu\right)^2}{2 \cdot \sigma^2}\right)}$$

Taking the logarithm of $P(X|\theta)$

$$= \ln\left(\left(\frac{1}{\left(\sigma \cdot \sqrt{2\pi}\right)}\right)^N\right) + \sum_{i \in \{1,..N\}} -\frac{\left(x_i-\mu\right)^2}{2 \cdot \sigma^2}$$ ----------------------------(I)

Differentiating (I) wrt μ,

$$\frac{\delta}{\delta\mu} \rightarrow \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N\right) - 0.5 \sum_{i \in \{1,..N\}} \frac{\left(x_i-\mu\right)^2}{\sigma^2} = 0$$

$$0 - 0.5 \sum_{i \in \{1,..N\}} \frac{1}{\sigma^2} \cdot 2 \cdot \left(x_i - \mu^*\right) \cdot (-1) = 0$$

$$\sum_{i \in \{1,..N\}} \frac{1}{\sigma^2} \cdot \left(x_i - \mu^* \right) = 0$$

$$\sum_{i \in \{1,..N\}} \frac{x_i}{\sigma^2} - \sum_{j \in \{1,..N\}} \frac{\mu^*}{\sigma^2} = 0$$

$$\sum_{i = \{1,..N\}} \frac{x_i}{\sigma^2} = \sum_{j = \{1,..N\}} \frac{\mu^*}{\sigma^2}$$

$$\sum_{i \in \{1,..N\}} x_i = N \cdot \mu^*$$

$$\frac{1}{N} \sum_{\mathrm{i} \in \{1,..\mathrm{N}\}} \mathrm{x_i} = \frac{1}{N} \cdot N \cdot \overline{x} = \overline{x} - - - - - - - (1)$$

Differentiating (I) wrt σ,

$$\frac{\delta}{\delta\sigma} \rightarrow ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^N - \sum_{i \in \{1,..N\}} \frac{\left(x_i - \mu \right)^2}{2 \cdot \sigma^2} = 0$$

$$\frac{\delta}{\delta\sigma} \rightarrow ln(1)^N - ln\left(\sqrt{2\pi\sigma^2} \right)^N - \sum_{i \in \{1,..N\}} \frac{\left(x_i - \mu \right)^2}{2 \cdot \sigma^2} = 0$$

$$\frac{\delta}{\delta\sigma} \rightarrow 0 - ln\left(\sqrt{2\pi\sigma^2} \right)^N - \frac{\left(x_i - \mu \right)^2}{2 \cdot \sigma^2} = 0$$

$$\frac{\delta}{\delta\sigma} \rightarrow -ln\left(2\pi\sigma^2 \right)^{\frac{N}{2}} - \sum_{i \in \{1,..N\}} \frac{\left(x_i - \mu \right)^2}{2 \cdot \sigma^2} = 0$$

$$\frac{\delta}{\delta\sigma} \rightarrow -\frac{N}{2} ln\left(2\pi\sigma^2 \right) - \sum_{i \in \{1,..N\}} \frac{\left(x_i - \mu \right)^2}{2 \cdot \sigma^2} = 0$$

$$-\frac{N}{2} \frac{\delta}{\delta\sigma}\left(ln\left(2\pi\sigma^2 \right) \right) - \sum_{i \in \{1,..N\}} \frac{\left(x_i - \mu \right)^2}{2} \frac{\delta}{\delta\sigma}\left(\sigma^{-2} \right) = 0$$

$$-\frac{N}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi \cdot \frac{\delta}{\delta\sigma}\left(\sigma^2 \right) - \sum_{i \in \{1,..N\}} \frac{\left(x_i - \mu \right)^2}{2} \left(-2 \cdot \sigma^{-3} \right) = 0$$

$$-\frac{N}{2}\cdot\frac{1}{2\pi\sigma^2}\cdot(4\pi\sigma)\ -\sum_{i=\{1,..N\}}\frac{\left(x_i-\mu\right)^2}{2}\left(\frac{-2}{\sigma^3}\right)=0$$

$$-\frac{N}{\sigma}+\frac{1}{\sigma^3}\ \cdot\sum_{i=\{1,..N\}}\left(x_i-\mu\right)^2\ =0$$

$$\frac{1}{\sigma^3}\sum_{i\in\{1,..N\}}\left(x_i-\mu\right)^2=\frac{N}{\sigma}$$

$$\sum_{i\in\{1,..N\}}\left(x_i-\mu\right)^2=N\cdot\sigma^2$$

$$\sum_{i\in\{1,..N\}}\left(x_i-\overline{x}\right)^2=N\cdot\sigma^2 \qquad \{from\ (1)\}$$

$$\boxed{Sum-Of-Squared-Error\ Loss=\sum_{i\in\{1,..N\}}\left(x_i-\overline{x}\right)^2}$$