# CS4801: Principle of Machine Learning
# Assignment 3

**deadline : 3rd Ocotber 2018 morning 8am**

This homework consists of problems covering Classification. A few instructions to make life easier for all of us:

- Assignment need not to be submitted.
- Quiz on this Assignment will be held on 3rd Ocotber, during class hours.

| age | income | software_engineer | credit_rating | buys_computer |
|---|---|---|---|---|
| <=20 | high | no | fair | no |
| <=20 | high | no | excellent | no |
| 21-30 | high | no | fair | yes |
| >30 | medium | no | fair | yes |
| >30 | low | yes | fair | yes |
| >30 | low | yes | excellent | no |
| 21-30 | low | yes | excellent | yes |
| <=20 | medium | no | fair | no |
| <=20 | low | yes | fair | yes |
| >30 | medium | yes | fair | yes |
| <=20 | medium | yes | excellent | yes |
| 21-30 | medium | no | excellent | yes |
| 21-30 | high | yes | fair | yes |
| >30 | medium | no | excellent | no |

Figure 1: Decision Tree

## Exercise 1: Classification

(a) (2 points each)Please go through at-least 10 data set for classification from UC Machine Learning Repository ( `https://archive.ics.uci.edu/ml/datasets.html?format=&task=cla&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=table`) and Study what is feature and what is output for classifier.

(b) (2 points)Design $K$ class classifier with help of 2 class SVM.

(c) (5 points)Consider the following data set in Figure 1 and build a DT with your choice impurity measure and find the prediction for

$X$ = (age=20, income = medium, software engineer = yes, credit rating = fair).

Please state why you have considered that choice of impurity measure?

(d) (1 points) Consider two non-negative numbers $a$ and $b$ and show that, if $a \leq b$, then $a \leq (ab)^{\frac{1}{2}}$. Use this result to show that, if the decision region of a two-class classification problem are chosen to minimize the probability of mis-classification, this probability will satisfy

$$p(mistake) \leq \int \{p(\mathbf{x}, \mathcal{C}_1)p(\mathbf{x}, \mathcal{C}_2)\}^{\frac{1}{2}} d\mathbf{x}$$

(e) (2 point) Derive the update equation for gradient descent approach to solve following problems

i. Logistic regression: $E(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} log(1 + e^{-y_i(\mathbf{x}_i^T\mathbf{w})}) + \frac{\lambda}{2}\|\mathbf{w}\|_2^2$

(f) (2 points) The error of a binary classifier which guesses completely randomly is 0.5. What is the error of a random k-class classifier for $k > 2$ labels.

i. Random guesser $G$ knows that there are $k$ labels, and for each example, selects a label out of $\{1, ..., k\}$ uniformly at random. What is the error of $G$ ?

ii. Now suppose we have a more sophisticated random guesser $Z$ who knows that $w_1$ fraction of the data distribution has label 1, $w_2$ fraction has label 2, and so on. For each example, Z also selects a label out of $\{1, ..., k\}$ at random, but he selects label 1 with probability $w_1$, label 2 with probability $w_2$ and so on. What is the error of Z?

(g) (2 points) Consider the following two data distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ over labeled examples. There is a single feature, denoted by $X$ which takes values in the set $\{1, 2, 3, 4\}$ and a binary label $Y \in \{0, 1\}$. $\mathcal{D}_1$ is described as follows:

$$\begin{aligned} Pr(X = i) &= \frac{1}{4}, \quad i \in \{1, 2, 3, 4\} \\ Pr(Y = 1|X = i) &= 1, \quad i \in \{1, 4\} \\ Pr(Y = 0|X = i) &= 1, \quad i \in \{2, 3\} \end{aligned} \tag{1}$$

$D_2$ is described as follows.

$$\begin{aligned} Pr(X = i) &= \frac{1}{4}, \quad i \in \{1, 2, 3, 4\} \\ Pr(Y = 1|X = i) &= \frac{i}{10}, \quad i \in \{1, 2, 34\} \end{aligned} \tag{2}$$

i. Consider the following classifier

$$h : h(x) = 1 \text{ if } x > 1.5 \text{ and } 0 \text{ otherwise.}$$

What is the true error of $h$ when the true data distribution is $D_1$?

ii. Suppose our classifier is

$$h_t : h(x) = 1 \text{ if } x > t \text{ and } 0 \text{ otherwise}$$

. Find $t$ which minimizes the true error of $h_t$ when the true data distribution is $D_1$. What is the true error of this classifier?

iii. Repeat parts **i.** and **ii.** for the data distribution D2.