Pattern Recognition

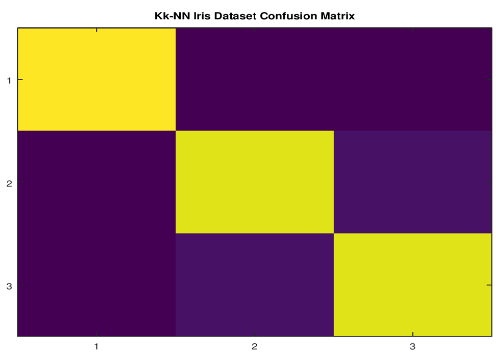
Exercise 3

Report Delivery Date: 3 June 2018

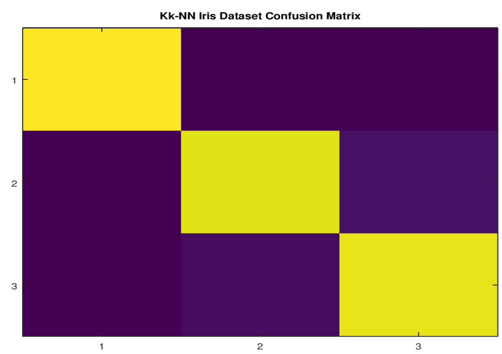
Student: Konidaris Vissarion 2011030123

Software: Matlab code

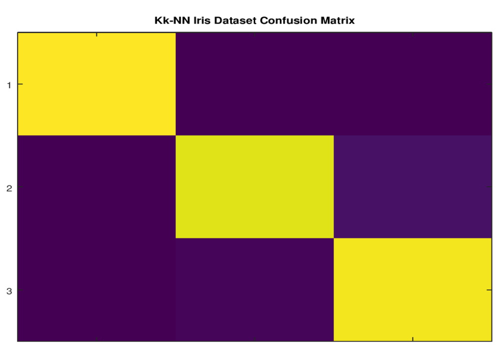
1. In this exercise we had to implement the k Nearest Neighbors algorithm for the classification of the Iris Dataset. The implementation of the kNN algorithm can be found in the $exercise3\_1$ directory. The classifiers' accuracy reaches 98 percent for $K = 21$. Below are the confusion matrices for some values of K.



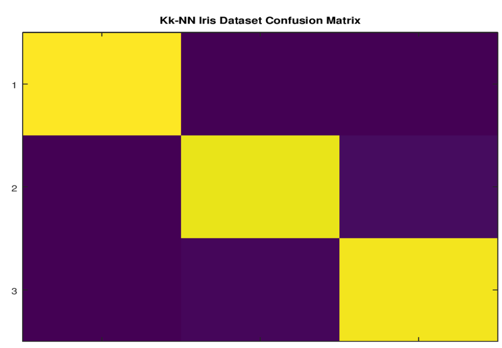
(a) $K = 3$, $accuracy : 0.96$



(b) $K = 5$, $accuracy : 0.96667$



(c) $K = 15$, $accuracy : 0.97333$



(d) $K = 21$, $accuracy : 0.98$

Figure 1: Confusion matrices of kNN classifier on the Iris dataset.

2. In this exercise we got familiar with the Logistic Regression classification algorithm. Suppose we have a set of m examples $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), ..., (x^{(m)}, y^{(m)})\}$, where $x^{(i)} \in \mathbb{R}$ and $y^{(i)} \in \{0, 1\}$. For a given x we want to predict its' true label y. The logistic regression hypothesis is defined as

$$h_\theta = g(\theta^T x)$$

where

$$g(z) = \frac{1}{1 + e^{-z}}.$$

If $\hat{y}^{(i)} = h_\theta(x^{(i)})$ is the estimation/hypothesis of the logistic regression for the true label $y^{(i)}$ then the loss function is defined as

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\Big(- y^{(i)} ln\Big(\hat{y}^{(i)}\Big) - (1 - \hat{y}^{(i)}) ln\Big(1 - \hat{y}^{(i)}\Big)\Big).$$

By substitution we get

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\Big(- y^{(i)} ln\Big(h_\theta(x^{(i)})\Big) - (1 - \hat{y}^{(i)}) ln\Big(1 - h_\theta(x^{(i)})\Big)\Big).$$

We know calculate the slope of the loss function $J(\theta)$ w.r.t the learnable parameters θ of our learner.

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

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

By substitution we get that

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\Big(y^{(i)} ln\Big(1 + e^{-\theta^T x^{(i)}}\Big) + (1 - y^{(i)})\Big(\theta^T x^{(i)} + ln(1 + e^{-\theta^T x^{(i)}})\Big)\Big) =$$

$$\frac{1}{m}\sum_{i=1}^{m}\Big(ln\Big(e^{\theta^T x^{(i)}}(1 + e^{-\theta^T x^{(i)}})\Big) - y^{(i)}\theta^T x^{(i)}\Big) =$$

$$= \frac{1}{m}\sum_{i=1}^{m}\Big(ln(1 + e^{\theta^T x^{(i)}}) - y^{(i)}\theta^T x^{(i)}\Big)$$

By taking the derivative of the above relation w.r.t the learnable parameters $\theta \in \mathbb{R}^n$ we get that

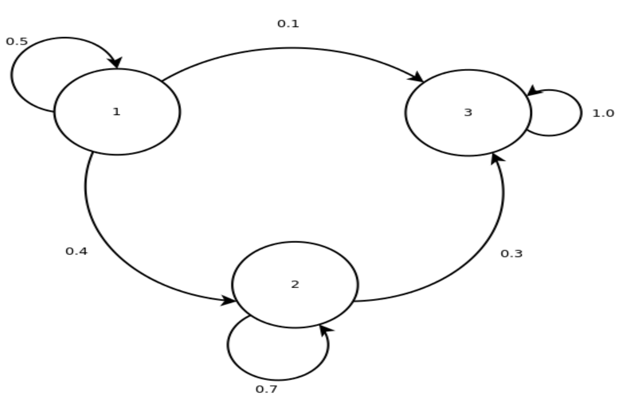
$$\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m}\sum_{i=1}^{m}\Big(\frac{x^{(i)}e^{\theta^T x^{(i)}}}{1+e^{\theta^T x^{(i)}}} - y^{(i)}x^{(i)}\Big) = \frac{1}{m}\sum_{i=1}^{m}(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)})x^{(i)}.$$

Proving that

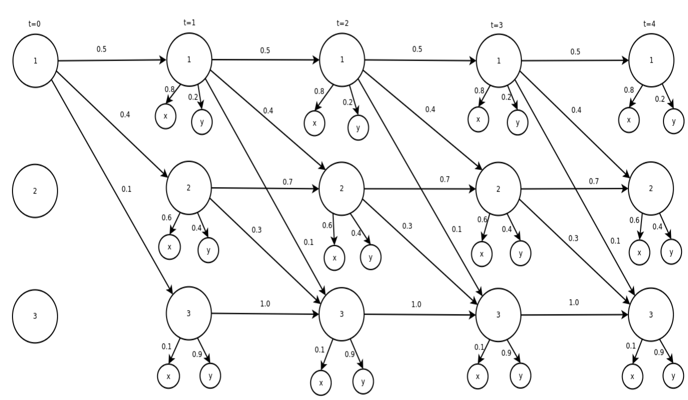
$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}.$$

The implemetation of the logistic regression classifier can be found in the $exercise3\_2$ directory. According to this implementation, for a student with scores 45 and 85, the classifier predicts an admission probability of 0.776289.

3. Suppose we have a Hidden Markov Model (HMM) λ with 3 states and two observations $[x, y]$. The initial state probabilities are $P(S_{t_1} = q_1) = 1$, $P(S_{t_1} = q_2) = 0$ and $P(S_{t_1} = q_3) = 0$. The transition probabilities can be seen below on the diagram of this particular HMM.



(a) HMM diagram.


(b) Trellis diagram of length 4.

We know calculate the probability of observing the sequence $O = [xxxy]$ given the HMM λ, from all possible trials that end up in state q_3 at time $T = 4$.

$$P(O|\lambda) = P(y|S_{t4} = q_3)P(x|S_{t3})P(x|S_{t2})P(x|S_{t1})$$

$$P(x|S_{t1}) = P(x|S_{t1} = q_1)P(S_{t1} = q_1) = 0.8$$

$$P(x|S_{t2}) = P(x|S_{t2} = q_1)P(S_{t2} = q_1|S_{t1} = q_1) + P(x|S_{t2} = q_2)P(S_{t2} = q_2|S_{t1} = q_1)+$$

$$+P(x|S_{t2} = q_3)P(S_{t2} = q_3|S_{t1} = q_1) = 0.8 * 0.5 + 0.6 * 0.4 + 0.1 * 0.1 = 0.65$$

Let

$$\alpha = P(S_{t3} = q_1|S_{t2} = q_1)P(S_{t2} = q_1) + P(S_{t3} = q_1|S_{t2} = q_2)P(S_{t2} = q_2)+$$

$$+P(S_{t3} = q_1|S_{t2} = q_3)P(S_{t2} = q_3) = 0.5 * 0.5 + 0 + 0 = 0.25$$

$$\beta = P(S_{t3} = q_2|S_{t2} = q_1)P(S_{t2} = q_1) + P(S_{t3} = q_2|S_{t2} = q_2)P(S_{t2} = q_2)+$$

$$+P(S_{t3} = q_2|S_{t2} = q_3)P(S_{t2} = q_3) = 0.4 * 0.5 + 0.7 * 0.4 + 0 = 0.48$$

$$\gamma = P(S_{t3} = q_3|S_{t2} = q_1)P(S_{t2} = q_1) + P(S_{t3} = q_3|S_{t2} = q_2)P(S_{t2} = q_2)+$$

$$P(S_{t3} = q_3|S_{t2} = q_3)P(S_{t2} = q_3) = 0.1 * 0.5 + 0.3 * 0.4 + 1. * 0.1 = 0.27$$

Then

$$P(x|S_{t3}) = P(x|S_{t3} = q_1)\alpha + P(x|S_{t3} = q_2)\beta + P(x|S_{t3} = q_3)\gamma =$$

$$= 0.8 * 0.25 + 0.6 * 0.48 + 0.1 * 0.27 = 0.515$$

$$P(y|S_{t4} = q_3) = P(y|S_{t4} = q_3)\Big(P(S_{t4} = q_3|S_{t3} = q_1)\alpha + P(S_{t4} = q_3|S_{t3} = q_2)\beta$$

$$+P(S_{t4} = q_3|S_{t3} = q_3)\gamma\Big) =$$

$$= 0.9 * (0.1 * 0.25 + 0.3 * 0.48 + 1. * 0.27) = 0.3951$$

Finally we get that

$$P(O|\lambda) = P(y|S_{t4} = q_3)P(x|S_{t3})P(x|S_{t2})P(x|S_{t1}) =$$

$$= 0.3951 * 0.515 * 0.65 * 0.8 = 0.10580778$$

Lastly, given the observations O, the most probable route is $S_{t1} = q_1$, $S_{t2} = q_1$, $S_{t3} = q_1$ and $S_{t4} = q_3$with probability of happening $P^* = P(S_{t1} = q_1)P(S_{t2} = q_1|S_{t1} = q_1)P(S_{t3} = q_1|S_{t2} = q_1)P(S_{t4} = q_3|S_{t3} = q_1) = 1. * 0.5 * 0.5 * 0.1 = 0.025.$

4. In this final exercise we had to implement the K-means clustering algorithm for the purpose of using it to compress a 2D image. The implementation of the K-means algorithm can be found in the *exercise*3_4 directory. Below are some plots of the original and compressed image for some numbers of centroids K.

(a) 2 Centroids

(b) 4 Centroids

(c) 8 Centroids

(d) 16 Centroids

Figure 3: Image compression using K-means clustering.