CSE512 Spring 2021 - Machine Learning - Homework 3

Your Name: Sagar Onkar Toshniwal

Solar ID: 113260061

NetID email address: sagaronkar.toshniwal@stonybrook.edu

Names of people whom you discussed the homework with: None but I have taken reference from medium, edureka, Google, Stackoverflow, etc.

1]

1.1.1] Bayes risk for a given cost L for some action a_i for any given class C_j is given by

$$R(a_i/x) = \sum_j L(a_i/c_j)\, P(c_j/x)$$

It can be written as

Case 1] $\Rightarrow R(Y=1/x) = L(Y=1/\hat{y}=1)\, P(\hat{y}=1/x) +$
$L(Y=1/\hat{y}=0)\, P(\hat{y}=0/x)$

~~Case 1]~~ In this case, the term $L(Y=1/\hat{y}=1)\, P(\hat{y}=1/x)$ is zero as the prediction is correct. So there is no Risk / loss for this term

Now the Eq^n becomes

$$\Rightarrow L(Y=1/\hat{y}=0)\, P(\hat{y}=0/x)$$

cost of false positive $(L(Y=1/\hat{y}=0))$ is denoted by α.
$P(\hat{y}=0/x)$ is the prior. of negative.

$$\Rightarrow \alpha(1-\eta(x))$$

Case 2] $R(Y=0/x) = L(Y=0/\hat{y}=1)\, P(\hat{y}=1/x) +$
$L(Y=0/\hat{y}=0)\, P(\hat{y}=0/x)$

Here $L(Y=0/\hat{y}=0)\, P(\hat{y}=0/x) = 0$ as the prediction is correct.

$\Rightarrow \quad L(y=0/\hat{y}=1)\, P(\hat{y}=1/x)$

$\Rightarrow \quad 1 \cdot \eta(x)$ ... from given.

$\Rightarrow \quad \eta(x)$

$\therefore$ From case 1 & case 2

optimal Bayes classifier has the below risk

$$\boxed{r^*(x) = \min\left(\eta(x),\ \alpha\,(1-\eta(x))\right)}$$

1.1.2]

Let Total risk for a data point x whose nearest neighbour is some 'z' is $r(x)$

$\because$ An error is made when a wrong decision is taken because label $(x) \neq$ label (z).

From Que 1.1.1]

$$r(x) = L(y=0/\hat{y}=1)\, P(\hat{y}=1/x)\, P(\hat{y}=0/z) \; +$$

$$L(y=1/\hat{y}=1)\, P(\hat{y}=0/x)\, P(\hat{y}=1/z)$$

$\because$ we know & by given in Que.

$$L(y=0/\hat{y}=1) = 1$$

$$L(y=1/\hat{y}=1) = \alpha$$

$\therefore$ Now the eq^n becomes

$$r(x) = \eta(x)\,(1-\eta(z)) + \alpha\,(1-\eta(x))\,\eta(z)$$

As $n \to \infty$, $x \to z$

$$\therefore\ r(x) = \eta(x)\,(1-\eta(x)) + \alpha\,\eta(x)\,(1-\eta(x))$$

$$r(x) = (1+\alpha)\,\eta(x)\,(1-\eta(x))$$

1.1.3]

From 1.1.2]

$$r(x) = (1+\alpha)\,\eta(x)\,(1-\eta(x))$$

$$r^*(x) = \min\left(\eta(x),\ \alpha(1-\eta(x))\right)$$

$$r^*(x) = \eta(x) \quad \text{if } \eta(x) < \alpha(1-\eta(x)) \quad -①$$

$$= \alpha(1-\eta(x)) \quad \eta(x) > \alpha(1-\eta(x)) \quad -②$$

from ①,

we conclude the lower bound of $r^*(x)$ is $\eta(x)$. $\Rightarrow$ $\boxed{r^*(x) = \eta(x)}$

$$\therefore\ r(x) = (1+\alpha)\,r^*(x)\,(1-r^*(x))$$

... lower bound.

Let $r(x) < (1+\alpha)\,r^*(x)\,(1-r^*(x))$ when $\quad$ — ③

$$\eta(x) > \alpha(1-\eta(x))$$

Substitute $r(x)$ with $(1+\alpha)\,\eta(x)\,(1-\eta(x))$ in eq^n ③

$$\cancel{(1+\alpha)}\,\eta(x)\,\cancel{(1-\eta(x))} < \cancel{(1+\alpha)}\,\alpha\,\cancel{(1-\eta(x))}\,(1-\alpha(1-\eta(x)))$$

$$\eta(x) < \alpha(1-\alpha(1-\eta(x))) \rightarrow 4$$

since

$$\eta(x) > \alpha(1-\eta(x))$$

$$\therefore -\eta(x) < -\alpha(1-\eta(x))$$

$$1-\eta(x) < 1-\alpha(1-\eta(x))$$

From Hint $\alpha > 1$, $1-\eta(x)$ & $\eta(x)$ will have a lower bound, thus on multiplying α, the other side will still remain dominant.

$$\therefore \eta(x) < \alpha(1-\alpha(1-\eta(x)))$$

$$\therefore \boxed{\eta(x) < \alpha(1-\alpha(1-\eta(x)))}$$ ... proved for 4

From 1, 2, 3 & 4

$$\boxed{r(x) \leq (1+\alpha)\, r^* x\, (1-r^*(x))}$$

1.1.4]

$$r(x) \leq (1+\alpha)\, r^* x\, (1-r^*(x))$$

Taking Expectation on both sides,

$$E[r(x)] \leq (1+\alpha)\, E[r^*(x)(1-r^*(x))]$$

$$\leq (1+\alpha)\left[E[r^*(x)] - E[r^*(x)^2]\right] \quad - 1$$

$$E[x^2] = E[x]^2 + Var[x]$$

Page 3

$$E[r^*(x)^2] \geq E[r^*(x)]^2$$

$$R \leq (1+\alpha)\ E[r^*(x)] - (1+\alpha)\ E[r^*(x)]$$ ... From (1)

$$R \leq (1+\alpha) R^* - (1+\alpha) R^{*2}$$

$$E[r^*(x)] = R^*$$

$$\boxed{R \leq (1+\alpha) R^* (1-R^*)}$$ ...

1.2.1] Let η be the probability of being positive.
Given than atleast $(K+1/2)$ points ~~of~~ out of K are +ve.

$\therefore$ Let $r(x)$ be the asymptotic risk for a point x.

$$r(x) = P\left(x = \text{Positive} / (K+1/2) \text{ points are not positive}\right) + P\left(x = \text{Negative} / (K+1/2) \text{ points are positive}\right)$$

$$= \eta(x)\left(1 - g(n,K)\right) + \left(1-\eta(x)\right)\left(g(n,K)\right)$$

$$= \eta(x) - \eta(x)\, g(n,K) + g(n,K) - \eta(x)\, g(n,K)$$

$$= \eta(x) + g(\eta, k) - 2\eta(x)\, g(\eta, k)$$

$$= \eta(x) + (1 - 2\eta(x))\, g(\eta, k)$$

$$\boxed{\gamma(x) \Rightarrow \eta(x) + (1 - 2(\eta(x)))\, g(\eta, k)}$$

1.2.2]

From 1.2.1

$$\gamma(x) = \eta(x) + (1 - 2\eta(x))\, g(\eta, k)$$

$\because \gamma^*(x) = \eta(x)$ if prob ≤ 0.5 i.e. $\eta(x) \leq 0.5$

$= 1 - \eta(x)$ $\eta(x) \geq 0.5$

when $\eta(x) \leq 0.5$, substitute $\cancel{\gamma^*(x)}$ $\eta(x)$ with $\gamma^*(x)$

$$\boxed{\gamma(x) = \gamma^*(x) + (1 - 2\gamma^*(x))\, g(\eta, k)} \quad -① $$

When $\eta(x) \geq 0.5$

now substitute $\gamma^*(x) \geq 0.5$

$$\Rightarrow (1 - \eta(x)) + (1 - 2(1 - \eta(x)))\, g(1 - \eta(x), k)$$

$$\Rightarrow g(1 - \eta(x), k) \Rightarrow 1 - g(\eta(x), k) \quad -②$$

from eq^n ②, sub $g(1 - \eta(x), k)$ as $1 - g(\eta(x), k)$

Page 4

$\Rightarrow 1-\eta(x)+1-g(n,K)-2+2\,g(\eta^{(x)},K)+$
$2\eta(x)-2\eta(x)\,g(\eta^{(x)},K)$

$\Rightarrow \eta(x)+g(\eta^{(x)},K)-2\eta(x)\,\cancel{g(\eta,K)}\,g(\eta(x),K)$

$$\boxed{=\eta(x)+g(\eta,K)(1-2\eta(x))} \quad \cdots ②$$

$$\boxed{r(x)\Rightarrow r^*(x)+g(r^{*}(x),K)(1-2r^*(x))} \quad \cdots ③$$

From, ①, ② & ③ ... Hence proved.

1.2.3

Prove :- $g(r^*(x),K)\leq \exp(-2(0.5-r^*(x))^2K)$

$\rightarrow$ From Hoeffding's Inequality,

$P(H(n)\leq(p-\varepsilon)n)\leq\exp(-2\varepsilon^2 n)$

$\therefore$

$P\left(H(K)\leq\frac{K-1}{2}\right)\leq\exp(\cancel{\quad}-2\varepsilon^2 n)$

$p\Rightarrow 1-r^*(x)$

Let us solve ε

$\frac{K-1}{2}=K(1-r^*(x)-\varepsilon)$

$\Longleftarrow \quad K-1 = 2K - 2K r^*(x) - 2K\varepsilon$

$\varepsilon = \frac{1}{2} - r^*(x)$

$\varepsilon \geq 0.5 - r^*(x)$

$\because \quad g(r^*(x), K) \leq \exp(-2\varepsilon^2 K)$

$$\boxed{g(r^*(x), K) \leq \exp\left(-2\,(0.5 - r^*(x))^2 K\right)}$$

...

1.2.4]

prove $\quad r(x) \leq r^*(x) + \frac{1}{\sqrt{2K}}$

we know that,

$$r(x) = r^*(x) + (1 - 2r^*(x))\, g(r^*(x), K)$$

$$g(r^*(x), K) = \frac{r(x) - r^*(x)}{1 - 2r^*(x)}$$

Also $\quad g(r^*(x), K) \leq \exp\left(-2\,(0.5 - r^*(x))^2 K\right)$

$$\frac{r(x) - r^*(x)}{1 - 2r^*(x)} \leq \exp\left(-2\,(0.5 - r^*(x))^2 K\right)$$

Let $\quad 1 - 2r^*(x) = \quad \Leftrightarrow -1$ to 1.

we know that $\quad r^*(x)$ has ~~min value~~

of range of 0 to 1

$\therefore 1 - 2r^*(x)$ has range of -1 to 1

$$\frac{\gamma(x) - \gamma^*(x)}{1 - 2\gamma^*(x)} \leq \exp\left(-2\left(\frac{1-2\gamma^*(x)}{2}\right)^2 k\right)$$

$$\frac{\gamma(x) - \gamma^*(x)}{(1)} \leq \exp\left(-\frac{1}{2}(1)^2 k\right)$$

Let us assume $1 - 2R^*(x) = 1$ considering correct prediction

$$\gamma(x) \leq \gamma^*(x) + \exp\left(-\frac{1}{2}k\right)$$

~~$\exp\left(\frac{1}{2}k\right)$ is less than $\frac{1}{\sqrt{2k}}$~~

$\therefore$

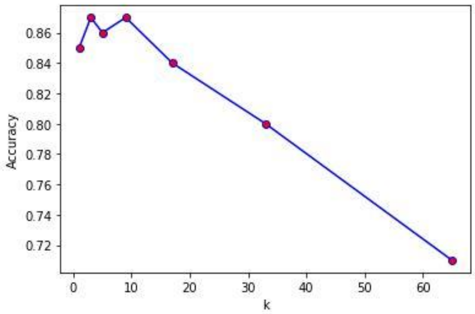
$$\boxed{\gamma(x) \leq \gamma^*(x) + \exp\left(-\frac{1}{2}k\right)}$$

$$\gamma(x) \leq \gamma^*(x) + \frac{1}{\sqrt{2k}}$$

Que.2.2.1]

```
[12]: m = [0, 1, 2, 4, 8, 16, 32]
      k_list = [x * 2 + 1 for x in m]
      acc_list = []
      for i in k_list :
          y_pred, idxs = knn_classifier(X_train,y_train,X_test, i)
          a = accuracy_score(y_test,y_pred)
          acc_list.append(a)
```

```
[80]: plt.plot (k_list,acc_list, marker='o', color='blue', mfc='red')
      plt.xlabel("k")
      plt.ylabel("Accuracy")
      plt.show()
```



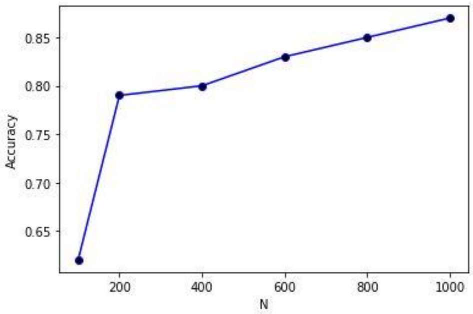
Yes k does affect the performance.

From the graph with the increase in k (post value of 10) there is a steep decrease in the accuracy. The model performs best at k = 3 and k = 9 with an accuracy of 87%.

Que 2.2.2]

```
[82]: n = [100,200,400,600,800,1000]
      acc = []
      for i in n :
          y_pred, idxs = knn_classifier(X_train[:i,:], y_train[:i], X_test, 3)
          a = accuracy_score(y_test,y_pred)
          acc.append(a)
```

```
[84]: plt.plot (n,acc, marker='o', color='b', mfc='black')
      plt.xlabel("N")
      plt.ylabel("Accuracy")
      plt.show()
```



Yes N (number of data points) does affect the performance.

From the Graph with the increase in number of Training data points the model is learning better and it can be reflected from its accuracy which is increasing with the increasing in the number of Training data points.

Que 2.2.3]

```
[91]: y_pred = knn_classifier1(X_train,y_train, X_test,3)
      y_pred2, Idxs = knn_classifier(X_train,y_train, X_test,3)
      print ("Accuracy score for Manhattan distance : ", accuracy_score(y_test,y_pred))
      print ("Accuracy score for Eucledean distance : ", accuracy_score(y_test,y_pred2))

      Accuracy score for Manhattan distance :  0.83
      Accuracy score for Eucledean distance :  0.87
```

No, Using Manhattan distance brings down the accuracy from 87% to 83%.

With the use of Euclidean distance, The model performs better.

Accuracy :-

Euclidean -> 87%

Manhattan -> 83%

Que 2.2.4]

```
neighbour id for Sample 1 is [373 919  91 792 133]

neighbour id for Sample 2 is [950 955 196 719 402]

neighbour id for Sample 3 is [150 674  37 542 464]
```

Incorrectly Predicted Label/Value in Xtest for Sample 1 is 1

Incorrectly Predicted Label/Value in Xtest for Sample 2 is 9

Incorrectly Predicted Label/Value in Xtest for Sample 3 is 7

```
neighbors for Sample 1
neighbors for Sample 2
neighbors for Sample 3
```