CSE512 Spring 2021 - Machine Learning - Homework 3

Name: Sabrina Margetic

Solar ID: 109898930

Net ID Email Address: sabrina.margetic@stonybrook.edu

Names of people whom you discussed the homework with: None (Only the TA)

1 Question 1: Nearest Neighbor Classifiers

1.1 1-NN with asymmetric loss

1.1.1

The optimal Bayes risk is the risk of what you forgo by making a decision.

Given:
cost of false negative = 1 (cost of predicting negative when positive)
cost of false positive = α (cost of predicting positive when negative)
$\eta(x)$ = prob that x is positive.
$1 - \eta(x)$ = prob that x is negative.

$\therefore$ risk of choosing positive = (prob of positive)(cost of false neg) = $(\eta(x))(1) = \eta(x)$
risk of choosing negative = (prob of negative)(cost of false positive) = $(1 - \eta(x))(\alpha)$

When you choose an option, you choose the option with the most gain, therefore, the risk that you forgo is the minimum risk of the two options.
$\therefore \boxed{r^*(x) = min\{\eta(x), \alpha(1 - \eta(x))\}}$

1.1.2

We can consider a data point x, in which the nearest point, z, is used for classification.

The risk is the risk associated with making a wrong decision.

risk of x being positive = (probability that x is positive)(probability that z is negative)(cost of false negative) = $\eta(x)(1 - \eta(z))$
risk of x being negative = (probability that x is negative)(probability that z is positive)(cost of false positive) = $(1 - \eta(x))(\eta(z))(\alpha)$

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

As n goes to infinity, z goes to x.
$\therefore r(x) = \eta(x)(1 - \eta(x)) + (1 - \eta(x))(\eta(x))(\alpha)$
$\boxed{r(x) = \eta(x)(1 - \eta(x))(1 + \alpha)}$

1.1.3

<u>Goal:</u> Prove that $r(x) \leq (1 + \alpha)r^*(x)(1 - r^*(x))$
From previous part: $r(x) = \eta(x)(1 - \eta(x))(1 + \alpha)$

Since $r^* = min\{\eta(x), \alpha(1 - \eta(x))\}$, we can take into account 2 scenarios: 1. when $r^*(x) = \eta(x)$ 2. $r^*(x) = 1 - \eta(x)$.

<u>1. When the risk of Optimal Bayes is $\eta(x)$:</u>

The inequality provided becomes: $\eta(x)(1 - \eta(x))(1 + \alpha) \leq (1 + \alpha)\eta(x)(1 - \eta(x))$
Crossing out like terms, we obtain: $1 \leq 1$, which is clearly true.

<u>1. When the risk of Optimal Bayes is $\alpha(1 - \eta(x))$:</u>

$\eta(x)(1-\eta(x))(1+\alpha) \leq (1+\alpha)\alpha(1-\eta(x))(1-\alpha(1-\eta(x)))$
Crossing out like terms, we obtain: $\eta(x) \leq \alpha(1-\alpha(1-\eta(x)))$
$\eta(x) \leq \alpha - \alpha^2(1-\eta(x)))$
Since the risk is associated with what is forgone, it stands to reason that when the risk is surely $\alpha(1-\eta(x))$, the value of this decision should be between 0 and .5. Similarly, the value of the decision that was chosen should be between .5 and 1. They are inversely proportional, so when one of these values goes up, the other goes down.
Therefore, we can see that the left hand side of this inequality is between the range of .5 and 1.
On the right hand side, we see a negative value of $\alpha^2(1-\eta(x)) = \alpha[\alpha(1-\eta(x))]$. α has a value greater than 1, but gets dampened by the fact that $\alpha(1-\eta(x))$ is between 0 and .5, it becomes only a fraction of itself. Therefore, the negative aspect of the right hand side is smaller than the positive α. This proves that the value on the right hand side will always be positive. But, the reason why the value on the right hand side will always be greater is the fact that $\eta(x)$ and $\alpha(1-\eta(x))$ are inversely proportional. As $\eta(x)$ gets larger, $\alpha(1-\eta(x))$ gets smaller, and therefore, the right hand side gets closer to the value of α, which is always greater than 1 (while the left hand side can only reach a max of 1). Similarly, as $\alpha(1-\eta(x))$ gets larger, with a max of .5, the right hand side gets closer to $.5\alpha$ which is always greater than .5. The left hand side inversely gets closer to .5 in response. Needless to say, the right hand side is greater.
$\boxed{\therefore r(x) \leq (1+\alpha)r^*(x)(1-r^*(x)) \text{ is proven}}$

1.1.4

Goal: prove $R \leq (1+\alpha)R^*(1-R^*)$

$R^* = E[r^*(x)]$
$R = E[r(x)] = E[\eta(x)(1-\eta(x))(1+\alpha)]$
Taking $r^*(x) = \eta(x)$, we have: $E[r(x)] = E[r^*(x)(1-r^*(x))(1+\alpha)]$
$= (1+\alpha)(E[r^*(x)] - E[r^*(x)^2])$
We can observe that $E[r^*(x)^2] = E[r^*(x)]^2 + var[r^*(x)] \geq E[r^*(x)]^2$
We can use this to alter the equation for R: $R \leq (1+\alpha)(E[r^*(x)] - E[r^*(x)]^2)$
$R \leq (1+\alpha)(R^* - R^{*2})$
$\therefore \boxed{R \leq (1+\alpha)R^*(1-R^*) \text{ is proven}}$

1.2 k-NN Classifier

1.2.1

1. Risk of x= positive, but gets classified as negative.
Probability that x = positive: $\eta(x)$
To get classified as negative, at most $\frac{k+1}{2} - 1$ points should equal positive.

To solve for this probability in terms of $g(\eta, k)$, the probability that $\frac{k+1}{2}$ points are positive, we can also write it as: $g(\eta, k) = \sum_{i=\frac{k+1}{2}}^{i=k} \binom{k}{i}\eta^i(1-\eta)^{k-i} = 1 - \sum_{i=0}^{i=\frac{k+1}{2}-1} \binom{k}{i}\eta^i(1-\eta)^{k-i}$
The probability that at most $\frac{k+1}{2} - 1$ points are positive $= \sum_{i=0}^{i=\frac{k+1}{2}-1} \binom{k}{i}\eta^i(1-\eta)^{k-i}$
$\therefore$ probability that there at most $\frac{k+1}{2} - 1$ positive points is the same as $1 - g(\eta, k)$
Since cost is assumed to be the same, the risk becomes: $= \eta(x)(1 - g(\eta(k), k))$

2. Risk of x= negative, but gets classified as positive.
Probability that x= negative: $1 - \eta(x)$

To get classified as positive, $g(\eta, k)$ is the probability that the $\frac{k+1}{2}$ of the surrounding points would be positive.
$= (1-\eta(x))g(\eta(k), k)$

As n goes to infinity, the probability of $\frac{k+1}{2}$ points getting classified as positive goes to the probability of x getting classified as positive.
$\boxed{\therefore r(x) = (1-\eta(x))g(\eta(x), k) + \eta(x)(1-g(\eta(x), k))}$

1.2.2

Prove that $r(x) = r^*(x) + (1-2r^*(x))g(r^*(x), k)$
Given that $r^*(x) = min\{\eta, 1-\eta\}$

We have 2 scenario's to prove: 1. when $r^*(x) = \eta$ 2. when $r^*(x) = 1-\eta$.

1. $r^*(x) = \eta(x)$:

Lets take a look at the LHS of what we are trying to prove: $r(x) = \eta(x)(1-g(\eta, k)) + (1-\eta)g(\eta, k)$ from the previous part.
$= \eta - \eta g(\eta, k) + g(\eta, k) - \eta g(\eta, k)$
$= \eta - 2\eta g(\eta, k) + g(\eta, k)$

Now, lets take a look at the RHS of what we are trying to prove:
$r^*(x) + (1-2r^*(x))g(r^*(x), k) = \eta + (1-2\eta)g(\eta, k)$
$= \eta + g(\eta, k) - 2\eta g(\eta, k)$

Since both sides are equivalent, the equality from $r^*(x) = \eta$, holds to be true.

1. $r^*(x) = 1-\eta(x)$:

Looking at the left hand side:
$r(x) = \eta(x)(1-g(\eta, k)) + (1-\eta)g(\eta, k)$
$= \eta - \eta g(\eta, k) + g(\eta, k) - \eta g(\eta, k)$
$= \eta - 2\eta g(\eta, k) + g(\eta, k)$
*these steps for the LHS were the same as that from the first part.

Looking at the RHS:
$r^*(x) + (1-2r^*(x))g(r^*(x), k) = 1-\eta + (1-2(1-\eta))g(1-\eta, k)$
$= 1-\eta + (-1+2\eta))(g(1-\eta, k))$
$= 1-\eta - g(1-\eta, k) + 2\eta g(1-\eta, k)$
From previously, $g(1-\eta, k) = 1-g(\eta, k)$
$\therefore 1-\eta - (g(1-\eta, k)) + 2\eta g(1-\eta, k) = 1-\eta-(1-g(\eta, k)) + 2\eta(1-g(\eta, k))$
$= 1-\eta-1+g(\eta, k)) + 2\eta - 2\eta g(\eta, k))$
$\eta + g(\eta, k) - 2\eta g(\eta, k)$

Since the RHS and the LHS equal, this equality holds true.
Since the equality holds true for both possible values of $r^*(x)$, it is thereby proven to be true.

1.2.3

Goal: Prove $g(r^*(x), k) \leq exp(-2k(.5 - r^*(x))^2)$ using Hoeffding's inequality.

Hoeffding Inequality: $Pr(S_n - E[S_n] \geq t) \leq exp(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2})$
Where $S_n = X_1 + ... + X_n$
and X_i is bounded by $[a_i, b_i]$

Since our problem is a bernoulli with only 2 outcomes, X_i is always bounded by $[0, 1]$

$E[S_n] = kr^*(x)$
$n = k$ since it is the number of times this event occurs.

By having $t = \frac{k+1}{2} - kr^*(x)$, the left hand side of the Hoeffding inequality becomes: $Pr(S_n - kr^*(x) \geq \frac{k+1}{2} - kr^*(x)) = Pr(S_n \geq \frac{k+1}{2}) = g(r^*(x), k)$.
Since $g(r^*(x), k)$ is the probability of $\frac{k+1}{2}$ data points getting classified as either negative or positive, depending on $r^*(x)$.

The right hand side becomes: $exp(\frac{-2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}) = exp(\frac{-2(\frac{k+1}{2}-kr^*(x))^2}{k})$
$= exp(\frac{-2(\frac{k+1}{2}-kr^*(x))(\frac{k+1}{2}-kr^*(x))}{k})$
$= exp(\frac{-2((\frac{k+1}{2})^2-2kr^*(x)\frac{k+1}{2}+k^2r^*(x)^2)}{k})$
$= exp(\frac{-2(\frac{1}{4}(k+1)^2-kr^*(x)(k+1)+k^2r^*(x)^2)}{k})$
$= exp(\frac{(\frac{-1}{2}(k+1)^2+2kr^*(x)(k+1)-2k^2r^*(x)^2)}{k})$
$= exp(\frac{(\frac{-1}{2}(k^2+2k+1)+2k^2r^*(x)+2kr^*(x)-2k^2r^*(x)^2)}{k})$
$= exp(\frac{(\frac{-1}{2}k^2-k-\frac{1}{2}+2k^2r^*(x)+2kr^*(x)-2k^2r^*(x)^2)}{k})$
$= exp(\frac{-1}{2}k - 1 - \frac{1}{2k} + 2kr^*(x) + 2r^*(x) - 2kr^*(x)^2)$
$= exp(k(\frac{-1}{2} + 2r^*(x) - 2r^*(x)^2) - 1 - \frac{1}{2k} + 2r^*(x))$
$= exp(k(\frac{-1}{2} + 2r^*(x) - 2r^*(x)^2))exp(-1 - \frac{1}{2k} + 2r^*(x))$
$= exp(-2k(\frac{1}{4} - r^*(x) + r^*(x)^2))exp(-1 - \frac{1}{2k} + 2r^*(x))$
$= exp(-2k(.5 - r^*(x))^2)exp(-1 - \frac{1}{2k} + 2r^*(x))$

Therefore, the original inequality holds true if:
$exp(-2k(.5 - r^*(x))^2)exp(-1 - \frac{1}{2k} + 2r^*(x)) \leq exp(-2k(.5 - r^*(x))^2)$
Which is the same as: $exp(-1 - \frac{1}{2k} + 2r^*(x)) \leq 1$

Now, lets look at the ranges of variables in this exponential.
<u>k:</u> k can range from 1 to infinity. So, the fraction: $\frac{1}{2k}$ ranges from .5 to 0.
<u>$r^*(x)$:</u> $r^*(x)$ can range from 0 to .5. Since $r^*(x) = min\{\eta, 1-\eta\}$, the probability that is forgone would be in the majority if greater than .5.

Therefore, the max of $exp(-1 - \frac{1}{2k} + 2r^*(x))$ would occur when $k = \infty$ and $r^*(x) = .5$, thereby making it equal to: $exp(-1 - 0 + 2(.5)) = exp(0) = 1$. The maximum value is 1. As k gets closer to 1, the value of the exponential gets more and more negative, thereby decreasing. Similarly, as $r^*(x)$ gets closer to 0, the positive $2r^*(x)$ value becomes less significant, and the negative values in the exponential take over.
$\therefore exp(-2k(.5 - r^*(x))^2)exp(-1 - \frac{1}{2k} + 2r^*(x)) \leq exp(-2k(.5 - r^*(x))^2)$
$\therefore g(r^*(x), k) \leq exp(-2k(.5 - r^*(x))^2)exp(-1 - \frac{1}{2k} + 2r^*(x)) \leq exp(-2k(.5 - r^*(x))^2)$
$\boxed{\therefore g(r^*(x), k) \leq exp(-2k(.5 - r^*(x))^2)}$

1.2.4

Prove that: $r(x) \leq r^*(x) + \frac{1}{\sqrt{2k}}$
From previous parts: $r(x) = r^*(x) + (1 - 2r^*(x))g(r^*(x), k)$ and $g(r^*(x), k) \leq exp(-2(.5 - r^*(x))^2k)$
If we plug in for $g(r^*(x), k)$, we obtain: $r(x) \leq r^*(x) + (1 - 2r^*(x))exp(-2(.5 - r^*(x))^2k)$
Comparing this to our original proof equation, we see that for the inequality to hold, the following condition must hold:
$(1 - 2r^*(x))exp(-2(.5 - r^*(x))^2k) \leq \frac{1}{\sqrt{2k}}$
Which is the same as: $(\sqrt{2k})(1 - 2r^*(x))exp(-2(.5 - r^*(x))^2k) \leq 1$

We can prove this by finding the maximum value of $(\sqrt{2k})(1 - 2r^*(x))exp(-2(.5 - r^*(x))^2k)$, which

occurs when it's derivative is 0.
let $\alpha = (\sqrt{2k})(1-2r^*(x))$
then $(\sqrt{2k})(1-2r^*(x))exp(-2(.5-r^*(x))^2k) = \alpha exp(-\frac{1}{4}\alpha^2)$
$\therefore \frac{\partial}{\partial\alpha}\alpha exp(-\frac{1}{4}\alpha^2) = exp(-\frac{1}{4}\alpha^2) + \alpha exp(-\frac{1}{4}\alpha^2)\frac{-1}{4}2\alpha$
$= exp(-\frac{1}{4}\alpha^2) + \frac{-1}{2}\alpha^2 exp(-\frac{1}{4}\alpha^2)$
$= exp(-\frac{1}{4}\alpha^2)[1 + \frac{-1}{2}\alpha^2]$
Setting this equal to 0:
$0 = exp(-\frac{1}{4}\alpha^2)[1 + \frac{-1}{2}\alpha^2]$
$0 = 1 + \frac{-1}{2}\alpha^2$
$-1 = \frac{-1}{2}\alpha^2$
$2 = \alpha^2$
$\alpha = +/-\sqrt{2}$
The equation: $\alpha exp(-\frac{1}{4}\alpha^2)$ is maximum when $\alpha = \sqrt{2}$ (and minimum when $\alpha = -\sqrt{2}$). This is the same as saying the equation: $(\sqrt{2k})(1-2r^*(x))exp(-2(.5-r^*(x))^2k)$ is maximum when $(\sqrt{2k})(1-2r^*(x)) = \sqrt{2}$.
If we plug back the value of $\sqrt{2}$ in for alpha, we see: $\sqrt{2}exp(-\frac{1}{4}\sqrt{2}^2) = \sqrt{2}exp(-\frac{1}{4}2) = \sqrt{2}exp(-\frac{1}{2}) =\approx .8578$
Since this maximum value is less than 1, we can say that this function is indeed bounded by 1, ie. $(\sqrt{2k})(1-2r^*(x))exp(-2(.5-r^*(x))^2k) \leq 1$ holds true.
$\therefore (1-2r^*(x))exp(-2(.5-r^*(x))^2k) \leq \frac{1}{\sqrt{2k}}$ holds true.
$\therefore r(x) \leq r^*(x) + (1-2r^*(x))exp(-2(.5-r^*(x))^2k) \leq r^*(x) + \frac{1}{\sqrt{2k}}$ is also true.
We have thereby proved that $r(x) \leq r^*(x) + \frac{1}{\sqrt{2k}}$

2 Implementation

2.1

2.2

2.2.1

In Fig. 1 we are able to see that in the beginning, within a certain threshold (in our example I would take this threshold to be between 1 and about 15), the accuracy is affected by the value of k, however, it is all within a similar range. After that threshold, we observe a steady decrease in accuracy as k increases.

2.2.2

In Fig. 2 we observe that as we increase the amount of training data, the accuracy improves. We also observe a steeper increase in accuracy when n values are small and begin to increase; accuracy seemingly plateaus with larger values of n.

2.2.3

Euclidean Distance Accuracy: 0.87 = 87%
Manhattan Distance Accuracy: 0.84 = 84%

In this case, it is preferable to use Euclidean distance rather than Manhattan distance since we observe a increase in accuracy. However, the difference is still only 3%, so it is also doesn't make an extreme difference to choose one over the other.

2.2.4

From Fig. 3 we observe 3 failed cases 1. when the predicted label is 1 and the actual label is 2, 2. when the predicted label is 9 and the actual label is 8, and 3. when the predicted label is 7 and the actual

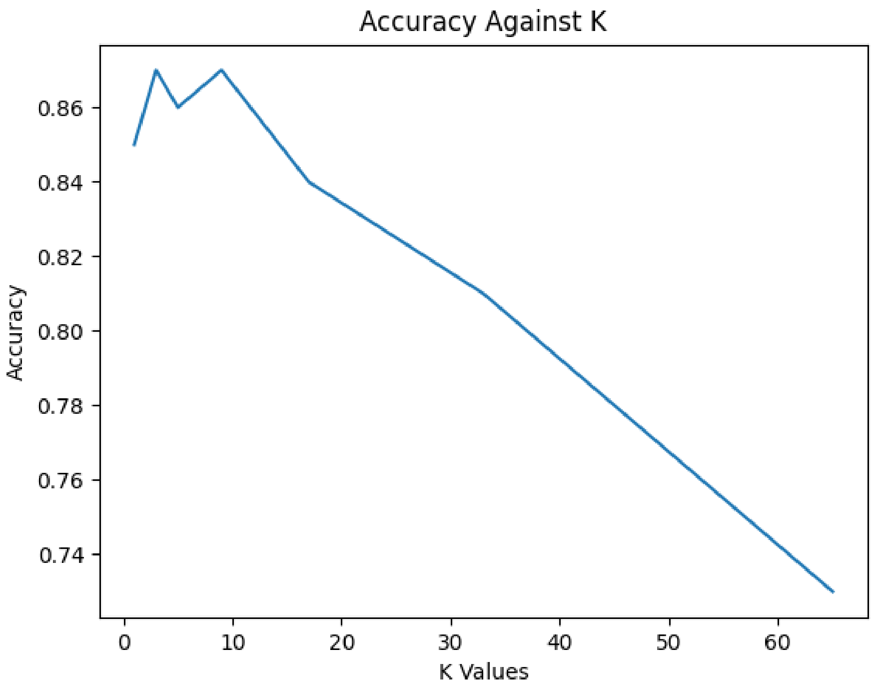

Figure 1

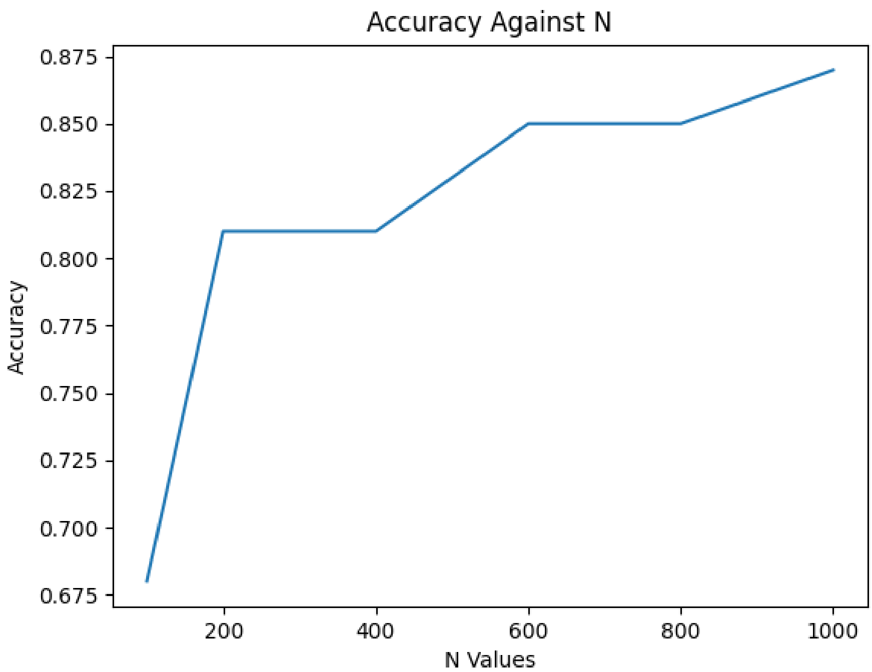

Figure 2: Accuracy against n, where n is the number of data points.

label is 2. In the first case we interestingly see that the vast majority of labels is 1, however, the closest nearest neighbor has a value of 2. In the second and third case, I believe the misclassification to be due to similar shapes. The 9 has a circular shape that is similar to and 8. The 7 with a dash passing through it resembles a 2.

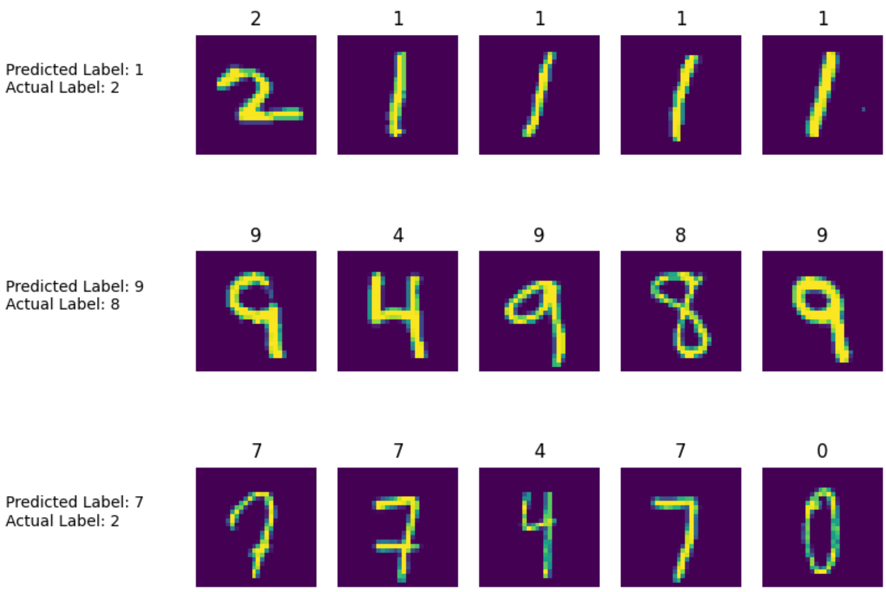
2
1
1
1
1
Predicted Label: 1
Actual Label: 2
9
4
9
8
9
Predicted Label: 9
Actual Label: 8
7
7
4
7
0
Predicted Label: 7
Actual Label: 2

Figure 3