Giang Duong – ID 014533857
Assignment 5 – CS 271
Professor Mark Stamp

Q1.

Giang Duong ID: 014533857
CS271 - HW5.

1) Use account 58 to solve the problem.

$$L = 16 - (x^2 + y^2) + \lambda(2x - y + 4) \quad (1)$$

take partial derivative and set it = 0.

$$\frac{\partial L}{\partial x} = -2x + \lambda(2) = 0. \Leftrightarrow x = \lambda$$

$$\frac{\partial L}{\partial y} = -2y + \lambda(-1) = 0 \Leftrightarrow y = \frac{-\lambda}{2}$$

Substituting while using the x, y value

$$L(\lambda) = 16 - \left(\lambda^2 + \left(\frac{+\lambda}{2}\right)^2\right) + \lambda\left(2\lambda + \frac{\lambda}{2} + 4\right)$$

$$L(\lambda) = 16 - \lambda^2 - \frac{\lambda^2}{4} + 2\lambda^2 + \frac{\lambda^2}{2} + 4\lambda.$$

$$L = 16 + \frac{5\lambda^2}{4} + 4\lambda \Leftrightarrow \cancel{\lambda =}$$

$$L = \frac{64 + 5\lambda^2 + 16\lambda}{4}$$

$$\frac{\partial L}{\partial \lambda} = 10\lambda + 16 = 0. \Leftrightarrow \lambda = -16/10 = -8/5$$

$\Leftrightarrow$ $x = -8/5$
$y = 4/5$
$\lambda = 8/5$

To validate: $g(1) = 2x - y + 4 = 2\left(\frac{-8}{5}\right) - \left(\frac{4}{5}\right) + 4 = 0.$ $\boxed{\checkmark}$

Q3/

(3). $p_i = 2^{-(\lambda + 1/\ln 2)} = \frac{1}{e} 2^{-\lambda}$ for $i = 1, 2, \dots n$. (5.10)

and $\sum_{i=1}^{n} p_i = 1$ (5.11)

Show that $p_1 = p_2 = \cdots p_n = \frac{1}{n}$.

We have n values.

$\sum_{i=1}^{n} pi = 1 = n \cdot \frac{1}{e} 2^{-\lambda}$. Note: $\left(\frac{1}{n} \cdot n = 1\right)$.

take log both side

$\log(1) = \log\left(n \cdot \frac{1}{e} 2^{-\lambda}\right)$

$\Leftrightarrow \quad 0 = \log \frac{n}{e} + \log 2^{-\lambda}$.

$\Leftrightarrow \quad 0 = \log \frac{n}{e} + (-\lambda) \log 2$.

$\Leftrightarrow \quad 0 = \log \frac{n}{e} - \log 2 . \lambda \quad \Leftrightarrow \quad \lambda = \frac{\log \frac{n}{e}}{\log 2}$

$\Leftrightarrow \quad \lambda = \log_2 \left(\frac{n}{e}\right) \quad \left[\log_a N = \frac{\log N}{\log a}\right]$

We have $p_i = \frac{1}{e} 2^{-\lambda} = \frac{1}{e} 2^{-\log_2\left(\frac{n}{e}\right)} = \frac{1}{e \, 2^{\log_2 n/e}}$

$\Leftrightarrow \quad \frac{1}{e \cdot n/e} = \frac{1}{n}$. $\boxed{\checkmark}$

Q4/

④. $L(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j z_i z_j (X_i \cdot X_j)$ (5.20)

where $X_i \cdot X_j = (x_i, y_i)\cdot(x_j, y_j) = x_i x_j + y_i y_j$
(dot product)

$L(w_1, w_2, b, \lambda) = \frac{w_1^2 + w_2^2}{2} + \sum_{i=1}^{n} \lambda_i (1 - z_i(w_1 x_i + w_2 y_i + b))$

$\frac{\partial L}{\partial w_1} = \frac{2w_1}{2} - \sum_{i=1}^{n} \lambda_i z_i x_i = 0 = w_1 = \sum_{i=1}^{n} \lambda_i z_i x_i$

$\frac{\partial L}{\partial w_2} = \frac{2w_2}{2} - \sum_{i=1}^{n} \lambda_i z_i y_i = 0 = w_2 = \sum_{j=1}^{n} \lambda_i z_i y_i$

$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} \lambda_i z_i = 0$

$\frac{\partial L}{\partial \lambda_i} = 1 - z_i(w_1 x_i + w_2 y_i + b) = 0$ for all $i = 1, 2, 3 \ldots n$

$\cancel{L(\lambda)} =$

$L(w_1, w_2, b, \lambda) = \frac{\left(\sum_{i=1}^{n} \lambda_i z_i x_i\right)^2 + \left(\sum_{j=1}^{n} \lambda_i z_i y_i\right)^2}{2} + \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \lambda_i z_i (w_1 x_i + w_2 y_i + b)$

we have $\sum \lambda_i z_i = 0$.

$\Rightarrow L(w_1, w_2, b, \lambda) = \frac{\left(\sum_{i=1}^{n} \lambda_i x_i z_i\right)^2 + \left(\sum_{i=1}^{n} \lambda_i z_i y_i\right)^2}{} + \sum_{i=1}^{n} \lambda_i (1 - 0)$

$\Leftrightarrow L(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} (\lambda_i x_i z_i)(\lambda_j x_j z_j)\right) + \left(\sum_{i=1}\sum_{j=1} (\lambda_i z_i y_i)(\lambda_j z_j y_j)\right)$.

$= L(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} (\lambda_i z_i \lambda_j z_j)(x_i x_j + y_i y_j)\right)$.

$L(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j z_i z_j (X_i \cdot X_j)\right)$ where
$(X_i X_j) = x_i x_j + y_i y_j$ $\boxed{\checkmark}$

15. If MIPS instruction format would have ... what would be max number of instruction be there for MIPS ...

29/

9) $f(X) = \sum_{i=1}^{n} \lambda_i z_i (X_i X) + b$. (5.21)

a) determine the weight associated with each component of a scored vector.

We have (5.21) is rewrite of 5.20.

which is $f(X) = w_1 x + w_2 y + b = w \cdot x + b$.

Consider (5.21) $f(x) = \sum_{i=1}^{n} \lambda_i z_i (x_i . x) + b$. ...as $X_i . X = x_i x + y_i y$

$$= \sum_{i=1}^{n} \lambda_i z_i (x_i x + y_i y) + b$$

$\neq \sum_{i=1}^{n}$ 2 eqs point $x = (x, y) \Leftrightarrow \sum_{i=1}^{n} \lambda_i z_i (x_i + y_i) = w$.

b) When we train the SVM for determine the weights of each feature, The weight states the more importante of ~~combination of the features~~ each feature and hence more weight, more important, Apply this we can eliminate the lower weight of each features when compared to others and can achive dimensionality reduction.

c) PCA: we are reduced noises and determine the important in terms of Principal Component which uses the linear combination of features. We will not get the direction values of importante (weight) of the features and so it is difficult to say which features are important. Because of that, it is difficult to archive dimensionality reduction in terms of features. The eigenvectors and eigenvalues can be determined from calculated the covariance matrix (higher the values, higher the variances, more important feature) and then there will no required for training

SVM: the weights of the model are determined after performing the training
Pros: when we get the weight of each features so we can eliminate the features based on it's importaint weight (the more weight, the more important).

.b.

When we train the SVM for determine the weight of each feature. The weight states the more important of each eatures and hence the more it weights, the more it is important. Apply in this, we can eliminate the lower weight of each eatures when compared to others and can achive the dimensionality reducetion.

.c.

PCA:

Vhile working with PCA, we are reduced noises and determine the important in terms of Principal Component which uses he linear combination of features. We will not get the direct values of importance weight of the featurs and so it will be ard to say which features are important. Because of that, it will be difficult to archive dimesionality reduction in terms of eatures.

he eigenvectors and eigenvalues can be calculated from Covariance matrix (higher the value, higher the variance, more mportant features) and then there will no requires for trainning.

SVM:
Cons:
The weights of the model are determined after performing the trainning.
Pros:
When we get the weight of each features then we can eleminate the feature based on its weights.

10.a.
C = 1.0, p= 2
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9743589743589743

10.b.
C = 3.0, p= 2
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9743589743589743

10.c.
C = 1.0, p= 4
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 1 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9230769230769231

10.d.
C = 3.0, p= 4
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 1 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9230769230769231

11./
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0
1 0]
0.9487179487179487
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0
1 0]
0.9487179487179487
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0
1 0]
0.9487179487179487
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0
1 0]
0.9230769230769231
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
0.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0

1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 0]
.9743589743589743

```
12./
```

With all 3 features
.9487179487179487
Respective Weights of HMM, SSD, OGS
[0.19688024 -0.53494025 -0.70487142]]
The most important weight is OGS score when the least important weight is HMM score
Accuracy after removing the HMM feature ->
.9487179487179487
Respective Weights of SSD, OGS
[-2.44489465 -2.24186897]]
Accuracy after removing the HMM and OGS feature ->
.9487179487179487
Weight of SSD
[-3.87596913]]

15.a
0.   0.   2.5 0.   1.25 1.25]
5.999999999999998
[0] = 1.4999999999999982
[0] = 1
[1] = 2.7499999999999982
[1] = 1
[2] = 0.2500000000000018
[2] = 1
[3] = -3.499999999999999
[3] = -1
[4] = -1.0
[4] = -1
[5] = -1.0

[5] = -1

15.b

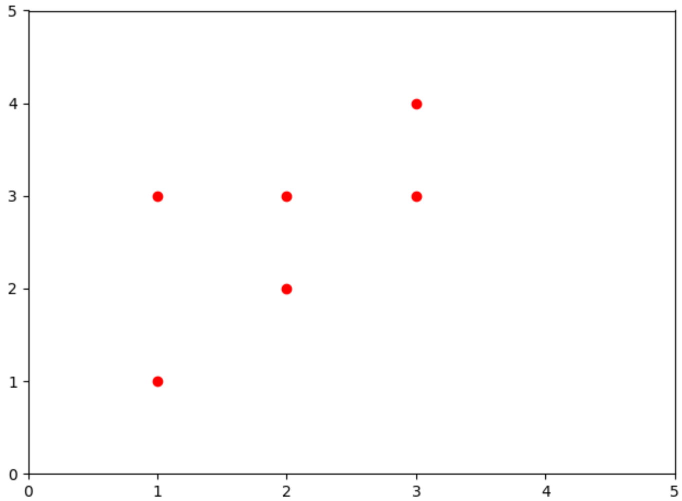
1.32173539 0.   3.09731961 0.   1.91905501 2.5   ]

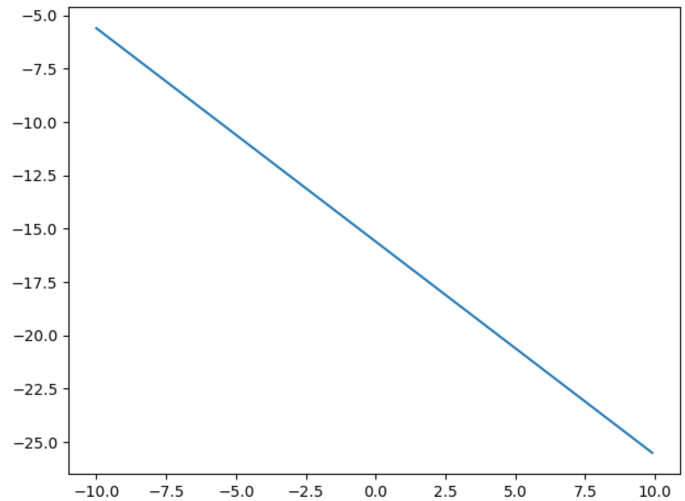
15.60197599369006

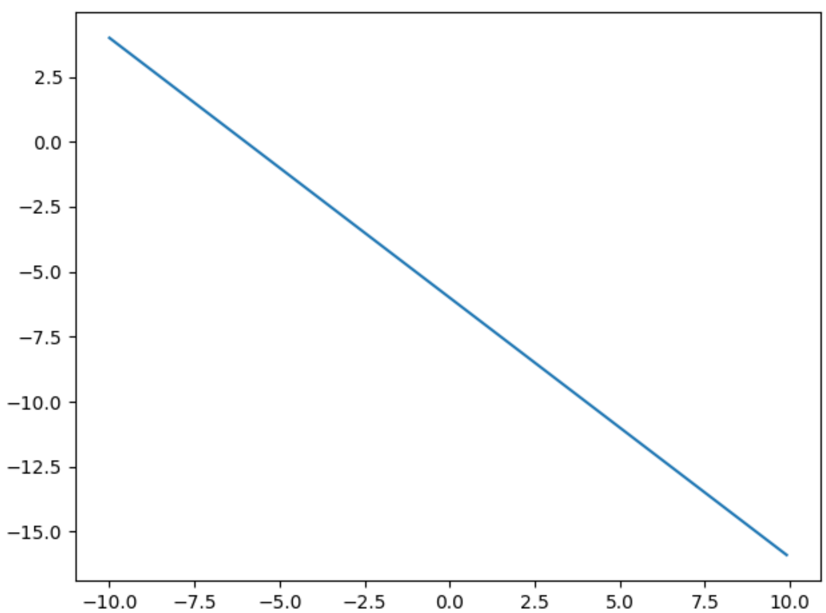
5.c.

Part a, we have b = -5.999999999999998 so the eqs for hyperplane is x + y =   -5.999999999999998

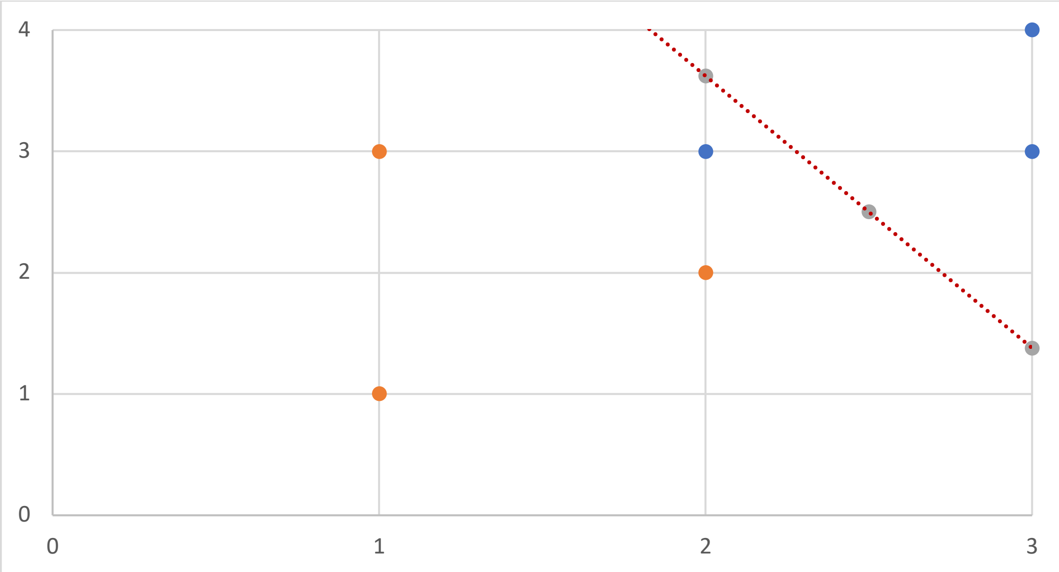
Part b, we have b = -15.60197599369006 so the eqs for hyperplane is x + y =   -15.60197599369006

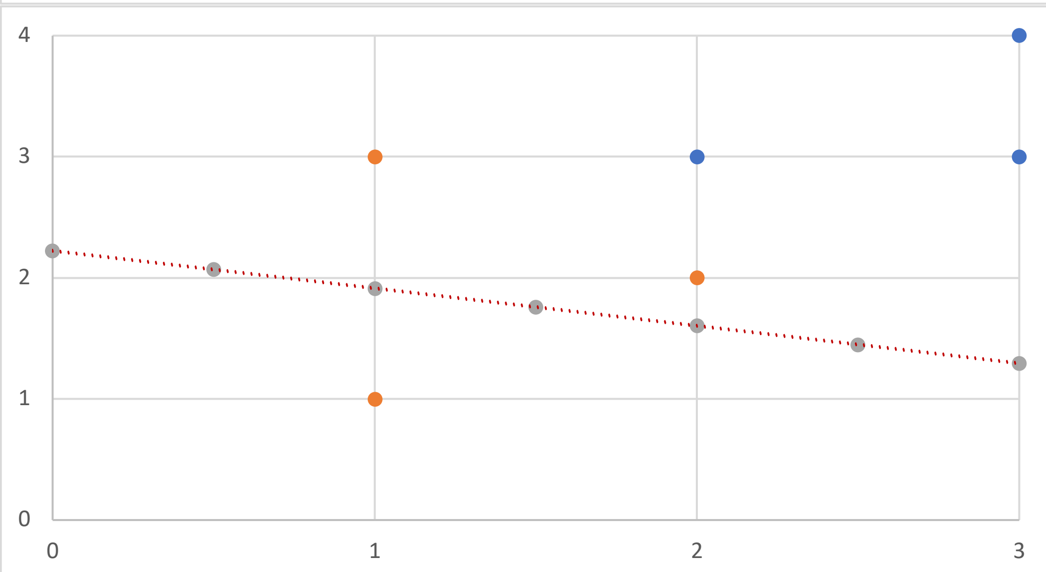






group	x	y
A	3	3
A	3	4
A	2	3
B	1	1
B	1	3
B	2	2





REFERENCES

1. M. Stamp, "Introduction to Machine Learning with Applications in Information Security."
2. Steven A.Cohen and Matthew W.Granade, "Models will run the world"