Homework 5

Nikhil Lahoti

012448514

Answer 1)

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

$\therefore \frac{dL}{dx} = -2x + \lambda(2) = 0. \quad \therefore x = \lambda$

$\frac{dL}{dy} = -2y + \lambda(-1) = 0. \quad \therefore y = -\lambda/2$

Substituting the value of x & y in eq^n (1)

$\therefore L(\lambda) = 16 - ((\lambda)^2 + (\lambda/2)^2) + \lambda(2(\lambda) - (-\lambda/2) + 4)$

Solving the equation further, we get.

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

taking derivative w.r.to λ we get.

$\frac{dL}{d\lambda} = \frac{10\lambda}{4} + 16 = 0.$

$\therefore 10\lambda = -16 \quad \therefore \lambda = -16/10 = -8/5$

$\therefore$ we get $x = \lambda = -8/5$

$y = -\lambda/2 = -(-8/5)/2 = 4/5$

$\therefore$ validating $\therefore$ $g(x) = (2x - y + 4)$

$= 2\left(-\frac{8}{5}\right) - \left(\frac{4}{5}\right) + 4$

$= -\frac{16}{5} - \frac{4}{5} + 4 = -\frac{20}{5} + 4 = 0.$

hence validated.

Answer 3)

Ans 3

We have $p_i = \frac{1}{e} 2^{-\lambda}$

$\therefore$ for n such values.

$$\sum_{i=1}^{n} p_i = n \frac{1}{e} 2^{-\lambda} = 1$$

$\therefore$ Taking log on both sides.

$$\log\left(\frac{n}{e} 2^{-\lambda}\right) = \log(1) = 0.$$

$$\therefore \log\left(\frac{n}{e}\right) + \log\left(2^{-\lambda}\right) = 0$$

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

$$\therefore \log\left(\frac{n}{e}\right) = \lambda \log 2$$

$$\therefore \lambda = \log\left(\frac{n}{e}\right) / \log 2 = \log_2\left(\frac{n}{e}\right)$$

$$\therefore p_i = \frac{1}{e} 2^{-\lambda}$$

$$= \frac{1}{e} 2^{-\log_2\left(\frac{n}{e}\right)} = \frac{1}{e\, 2^{\log_2\left(\frac{n}{e}\right)}}$$

$$= \frac{1}{\cancel{e} \times \left(\frac{n}{\cancel{e}}\right)} = \frac{1}{n}$$

hence proved.

Answer 4)

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

$$\frac{dL}{dw_1} = w_1 + \lambda_i(-z_i x_i) = 0$$

$$\therefore w_1 = \lambda_i z_i x_i \quad \text{-- (1)}$$

$$\frac{dL}{dw_2} = w_2 + \lambda_i(-z_i y_i) = 0 \quad \text{-- (2)}$$

$$\therefore w_2 = \lambda_i z_i y_i$$

$$\frac{dL}{db} = \sum_{i=1}^{n} \lambda_i z_i = 0 \quad \text{-- (3)}$$

$$\frac{dL}{d\lambda_i} = 1 - z_i(w_1 x_i + w_2 y_i + b) = 0 \quad \text{-- (4)}$$

We also have

$$w = \sum_{i=1}^{n} \lambda_i z_i x_i \quad \& \quad \lambda_i z_i = 0$$

Substituting these value in Lagrangian.

$$\therefore L(w_1, w_2, b, \lambda) =$$

$$\frac{\left(\sum_{i=1}^{n} \lambda_i z_i x_i\right)^2 + \left(\sum_{j=1}^{n} \lambda_j z_j x_j\right)^2}{2} + \sum_{i=1}^{n} \lambda_i (1 - 0)$$

as $w_1 = \sum_{i=1}^{n} \lambda_i z_i x_i$ & $w_2 = \sum_{j=1}^{n} \lambda_j z_j x_j$ &

$(1 - z_i(w_1 x_i + w_2 y_i + b) = 0$.

$\therefore$ We have.

$$L(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left[\left(\sum_{i=1}^{n} \lambda_i z_i x_i\right)^2 + \left(\sum_{i=1}^{n} \lambda_i z_i y_i\right)^2\right]$$

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

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

we can use $x_i x_j + y_i y_j = (x_i, y_i) \cdot (x_j, y_j) = (x_i \cdot x_j)$

$$\therefore = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} (\lambda_i \lambda_j z_i z_j (x_i \cdot x_j))\right]$$

hence proved.

Answer 6)

Q.6) We have equations as

Upper dashed Line

$w_1 x + w_2 y + b = +1$

$a_1 x + a_2 y = \alpha$

Lower dashed Line

$w_1 x + w_2 y + b = -1$

$a_1 x + a_2 y = \beta$

converting all the equations in y intercept form.

$y = -\frac{w_1 x}{w_2} + \frac{(1-b)}{w_2}$ ... (1)

$y = -\frac{a_1 x}{a_2} + \frac{\alpha}{a_2}$ ... (2)

$y = -\frac{w_1 x}{w_2} + \frac{(-1-b)}{w_2}$ ... (3)

$y = -\frac{a_1 x}{a_2} + \frac{\beta}{a_2}$ ... (4)

equating the c parts of equations [(1) & (2)] and [(3) & (4)]

$\therefore \frac{1-b}{w_2} = \frac{\alpha}{a_2}$

$\therefore \frac{w_2}{a_2} = \frac{1-b}{\alpha}$ ... (5)

$-\frac{(b+1)}{w_2} = \frac{\beta}{a_2}$

$\therefore -\frac{(b+1)}{\beta} = \frac{w_2}{a_2}$

Equating $\frac{w_2}{a_2}$ from both sides.

$\frac{1-b}{\alpha} = -\frac{(b+1)}{\beta}$

$\frac{(b-1)}{\alpha} = \frac{b+1}{\beta}$

$$\boxed{b = -\frac{(\alpha+\beta)}{\alpha-\beta}}$$

From eq (5)

$\frac{w_2}{a_2} = \frac{1-b}{\alpha}$ $\therefore w_2 = \frac{(1-b)}{\alpha} \times a_2$

$$\therefore w_2 = \frac{1 + \frac{(\alpha+\beta)}{\alpha-\beta}}{\alpha} \cdot a_2$$

$$= \frac{\alpha - \beta + \alpha + \beta}{\alpha(\alpha-\beta)} \cdot a_2 = \frac{2\alpha}{\alpha(\alpha-\beta)} \cdot a_2 = \boxed{\frac{2a_2}{(\alpha-\beta)}}$$

we have $\frac{w_1}{w_2} = \frac{a_1}{a_2}$ $\because$ as slopes are the same for both the lines.

$$\therefore w_1 = \frac{a_1 \times w_2}{a_2} = \frac{a_1 \times 2a_2}{a_2 \times (\alpha-\beta)}$$

$$\therefore \boxed{w_1 = \frac{2a_1}{\alpha-\beta}}$$

Answer 9) a)

Answer9) a)

$$f(x) = \sum_{i=1}^{n} \lambda_i z_i (X_i \cdot x) + b \quad \ldots (5.21)$$

$f(x)$ can also be written as.

$$f(x) = w_1 x + w_2 y + b = W \cdot X + b \quad \ldots (5\text{-}20)$$

Consider equation 5-21.

$$\therefore f(x) = \sum_{i=1}^{n} \lambda_i z_i (x_i \cdot x) + b$$

$$= \sum_{i=1}^{n} \lambda_i z_i (x_i x + y_i y) + b \quad \ldots \text{as } X_i \cdot X = x_i x + y_i y$$

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

Comparing the two equations point $X = (x, y)$

$$\therefore \sum_{i=1}^{n} \lambda_i z_i (x_i + y_i) = W$$

$$\therefore \boxed{W = \lambda_i z_i (x_i + y_i)}$$

b) We can train the SVM to determine the weights of each feature. The weights tell the importance of each feature and hence more the weight, important is the feature. This way, we can ignore the features which have lower weight as they are not important as compared to others and can achieve dimensionality reduction.

c) In case of PCA, we get the importance in terms of Principal Components which is the linear combination of the features. We do not get the direct values of importance (weights) of the features and so it is difficult to say which features are important. Thus, it is difficult to achieve dimensionality reduction in terms of features (not in Principal Component terms).

In case of SVM, we get the weights (importance) of each feature and so it is easy to discard features based on the weights, as lower weights correspond to less importance.

In case of PCA, the eigen vectors and the eigen values can be determined easily from the Covariance matrix (as higher the values, higher is the variance, important is the feature) , and hence no training is required. But in case of SVM, the weights of the model are determined after performing the training. This is the disadvantage of SVM.

Answer 12) Code is in the file: ***question12.py***

a) The accuracy obtained is 94.87 % (0.948717). The respective weights of HMM, SSD, OGS are [[0.19688024 -0.53494025 -0.70487142]]

b) Here is the output:

   With all 3 features

   0.9487179487179487

   Respective Weights of HMM, SSD, OGS

   [[0.19688024 -0.53494025 -0.70487142]]

   Accuracy after removing the HMM feature ->

   0.9487179487179487

   Respective Weights of SSD, OGS

   [[-2.44489465 -2.24186897]]

   Accuracy after removing the OGS feature ->

   0.9487179487179487

   Weight of SSD

[[-3.87596913]]

Answer 15)

a) Please find the code in **Question15a.py**

Sample Output

[0.   0.   2.5 0.   1.25 1.25]

-5.999999999999998

F[0] = 1.4999999999999982

z[0] = 1

F[1] = 2.7499999999999982

z[1] = 1

F[2] = 0.2500000000000018

z[2] = 1

F[3] = -3.499999999999999

z[3] = -1

F[4] = -1.0

z[4] = -1

F[5] = -1.0

z[5] = -1

b) Code in file **Question15b.py**

The values vary as the value of j is chosen randomly, here is one such instance

[0.22347138 0.   2.5   -0.54746154 2.67600985 0.59492308]

-4.7040393846154025

c) Please find the code in Question15c.ipynb

Sample output :