



Machine Learning Assignment 4

Calculations

Alireza Dastmalchi Saei

Student ID: 993613026

University: University of Isfahan

Course: Machine Learning Course

January 25, 2024

1 Question 1

1.1 a

Having the given constraints and objective, the Lagrangian function can be written as follows:

$$L = L(w, b, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^m \xi_i^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i + b) - 1 + \xi_i]$$

where the Lagrange multipliers have to satisfy the positivity constraints:

$$\alpha_i \geq 0 \quad (i = 1, \dots, m)$$

1.2 b

To minimize the Lagrangian function, we need to take partial derivatives and set them equal to zero:

1. For w :

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^m \alpha_i y_i x_i = 0 \implies w = \sum_{i=1}^m \alpha_i y_i x_i$$

2. For b :

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0 \implies \sum_{i=1}^m \alpha_i y_i = 0$$

3. For ξ_i :

$$\frac{\partial L}{\partial \xi_i} = C \xi_i - \alpha_i = 0 \implies C \xi_i = \alpha_i$$

After obtaining these solutions, we substitute them back into the Lagrangian function to get the minimum value.

1.3 c

Dual Form

By using equations found in *PartA* and *PartB*, we have both minimization and maximization problem for making the equation less complex, we calculate the dual form only with lagrangian multipliers and maximize it:

$$\begin{aligned}
L = L(w, b, \xi, \alpha) &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (a_i y_i x_i)(a_j y_j x_j) + \frac{1}{2} \sum_{i=1}^m \frac{\alpha_i}{\xi_i} \cdot \xi_i^2 \\
&\quad - \frac{1}{2} \sum_{i=1}^m \alpha_i [y_i ((\sum_{j=1}^m \alpha_j y_j x_j)^T x_i + b) - 1 + \xi_i] \\
&= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j + \frac{1}{2} \sum_{i=1}^m \alpha_i \xi_i - (\sum_{i=1}^m \alpha_i y_i) b \\
&\quad + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i \xi_i \\
&= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2} \sum_{i=1}^m \alpha_i \xi_i \\
&= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2} \sum_{i=1}^m \frac{\alpha_i^2}{C}
\end{aligned}$$

So the dual form of the SVM formula given in problem is:

$$\begin{aligned}
max_a &= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2} \sum_{i=1}^m \frac{\alpha_i^2}{C} \\
s.t. : &\alpha_i \geq 0 \quad (i = 1, \dots, m) \\
&\sum_{i=1}^m \alpha_i y_i = 0
\end{aligned}$$

2 Question 2

2.1 Given Information

We know that:

$$\begin{aligned}\Pr(P_1 = T) &= 0.4 \\ \Pr(P_2 = T \mid P_1 = T) &= 0.8 \\ \Pr(P_2 = T \mid P_1 = F) &= 0.5 \\ \Pr(P_3 = T \mid P_2 = T) &= 0.2 \\ \Pr(P_3 = T \mid P_2 = F) &= 0.3 \\ \Pr(P_4 = T \mid P_1 = T) &= 0.8 \\ \Pr(P_4 = T \mid P_1 = F) &= 0.5\end{aligned}$$

2.2 Value of $\Pr(P_3 = F)$

For calculation of $\Pr(P_3 = F)$ we must consider hidden variables and apply Variable Elimination method:

$$\begin{aligned}\Pr(p_3 = F) &= \sum_{p_2} \sum_{p_1} \sum_{p_4} \Pr(P_1 = p_1, P_2 = p_2, P_3 = F, P_4 = p_4) \\ &= \sum_{P_2 \in T, F} \sum_{P_1 \in T, F} \sum_{P_4 \in T, F} \Pr(P_1) \cdot \Pr(P_2 \mid P_1) \cdot \Pr(\neg p_3 \mid P_2) \cdot \Pr(P_4 \mid P_2) \\ &= \sum_{P_2 \in T, F} \sum_{P_1 \in T, F} \Pr(P_1) \cdot \Pr(P_2 \mid P_1) \cdot \Pr(\neg p_3 \mid P_2) \sum_{P_4 \in T, F} \Pr(P_4 \mid P_2) \\ &= \sum_{P_2 \in T, F} \sum_{P_1 \in T, F} \Pr(P_1) \cdot \Pr(P_2 \mid P_1) \cdot \Pr(\neg p_3 \mid P_2) \cdot 1 \\ &= \sum_{P_2 \in T, F} \Pr(\neg p_3 \mid P_2) \sum_{P_1 \in T, F} \Pr(P_1) \cdot \Pr(P_2 \mid P_1) \\ &= \sum_{P_2 \in T, F} \Pr(\neg p_3 \mid P_2) \tau_1(P_2) \\ &= (0.8 \cdot 0.62) + (0.7 \cdot 0.38) \\ &= 0.762\end{aligned}$$

2.3 Value of $\Pr(P_2 = T \mid P_3 = F)$

We calculated the value for $\Pr(P_3 = F)$, now we can find the value of $\Pr(P_2 = T \mid P_3 = F)$ via the formula of conditional probabilities:

$$\Pr(P_2 = T \mid P_3 = F) = \frac{\Pr(P_2 = T, P_3 = F)}{\Pr(P_3 = F)} = \frac{0.496}{0.762} = 0.6509$$

The value of $\Pr(P_2 = T, P_3 = F)$ is:

$$\begin{aligned} \Pr(P_2 = T, P_3 = F) &= \sum_{p_1} \sum_{p_4} \Pr(P_1 = p_1, P_2 = T, P_3 = F, P_4 = p_4) \\ &= \sum_{P_1 \in T, F} \sum_{P_4 \in T, F} \Pr(P_1) \cdot \Pr(p_2 \mid P_1) \cdot \Pr(\neg p_3 \mid p_2) \cdot \Pr(P_4 \mid p_2) \\ &= \sum_{P_1 \in T, F} \Pr(P_1) \cdot \Pr(p_2 \mid P_1) \cdot \Pr(\neg p_3 \mid p_2) \cdot \sum_{P_4 \in T, F} \Pr(P_4 \mid p_2) \\ &= \sum_{P_1 \in T, F} \Pr(P_1) \cdot \Pr(p_2 \mid P_1) \cdot \Pr(\neg p_3 \mid p_2) \cdot 1 \\ &= (0.4 \cdot 0.8 \cdot 0.8) + (0.6 \cdot 0.5 \cdot 0.8) \\ &= 0.496 \end{aligned}$$

3 Question 3

3.1 Dataset

We have following dataset and we are going to reduce dimensions using Principal Component Analysis (PCA):

Feature	Sample 1	Sample 2	Sample 3	Sample 4
X1	4	8	13	7
X2	11	4	5	14

3.2 Calculate Covariance Matrix

In this step, we are going to calculate the covariance matrix for the dataset. The covariance matrix is symmetric and square, with variances on the diagonal and covariances off the diagonal.

First we need to calculate the mean value for X_1 and X_2 :

$$\begin{cases} \text{Mean}_{X_1} = \frac{4 + 8 + 13 + 7}{4} = 8 \\ \text{Mean}_{X_2} = \frac{11 + 4 + 5 + 14}{4} = 8.5 \end{cases}$$

Then variances of X_1 and X_2 will be:

$$\begin{cases} \text{Var}(X_1) = \frac{(4 - 8)^2 + (8 - 8)^2 + (13 - 8)^2 + (7 - 8)^2}{3} = 14 \\ \text{Var}(X_2) = \frac{(11 - 8.5)^2 + (4 - 8.5)^2 + (5 - 8.5)^2 + (14 - 8.5)^2}{3} = 23 \\ \text{Cov}(X_1, X_2) = \frac{(4 - 8) \cdot (11 - 8.5) + 0 + (13 - 8) \cdot (5 - 8.5) + (7 - 8) \cdot (14 - 8.5)}{3} = -11 \end{cases}$$

The Covariance Matrix is:

$$\text{Covariance Matrix} = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

3.3 Calculate Eigenvalues and Eigenvectors

From the definition of the eigenvector v corresponding to the eigenvalue λ , we have $Av = \lambda v$.

Then $Av - \lambda v = (A - \lambda I)v = 0$.

Equation has a nonzero solution if and only if $\det(A - \lambda I) = 0$.

$$\begin{aligned}\det(A - \lambda I) &= \begin{vmatrix} 14 - \lambda & -11 \\ -11 & 23 - \lambda \end{vmatrix} \\ &= \lambda^2 - 37\lambda + 201 \\ &= \left(\lambda + \frac{\sqrt{565} - 37}{2}\right) \cdot \left(\lambda - \frac{\sqrt{565} + 37}{2}\right) \\ &= 0\end{aligned}$$

Eigenvalues of the Covariance Matrix are: $\lambda_1 = \frac{-\sqrt{565} + 37}{2} = 6.6151$ and $\lambda_2 = \frac{\sqrt{565} + 37}{2} = 30.3849$.

Eigen vectors of the Covariance Matrix are:

$$v_1 = \begin{bmatrix} 0.8302 \\ 0.5573 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 0.5573 \\ -0.8302 \end{bmatrix}$$

3.4 Sort eigenvalues and their corresponding eigenvectors

In this case we have only 2 eigenvalues and eigenvectors, so the only thing we need to do is to choose the bigger eigenvalue and its corresponding eigenvector (λ_2, v_2).

3.5 Reduce Dimension

Now, we need to calculate the product of original data matrix with v_2 :

$$\begin{bmatrix} 4 & 8 & 13 & 7 \\ 11 & 4 & 5 & 14 \end{bmatrix}^T \begin{bmatrix} 0.5573 \\ -0.8302 \end{bmatrix} = [-6.903 \quad 1.1376 \quad 3.0939 \quad -7.7217]$$

3.6 Original and Reduced Data

$$\text{Original Data} = \begin{bmatrix} 4 & 8 & 13 & 7 \\ 11 & 4 & 5 & 14 \end{bmatrix}$$

$$\text{Reduced Data} = [-6.903 \quad 1.1376 \quad 3.0939 \quad -7.7217]$$