122040012

1.1. (Learning of Linear Regression, 25 points) Suppose we have training data:

$$\{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), \ldots, (\boldsymbol{x}_N, \boldsymbol{y}_N)\},$$

where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $\boldsymbol{y}_i \in \mathbb{R}^k$, $i = 1, 2, \ldots, N$.

i) (9 pts) Find the closed-form solution of the following problem.

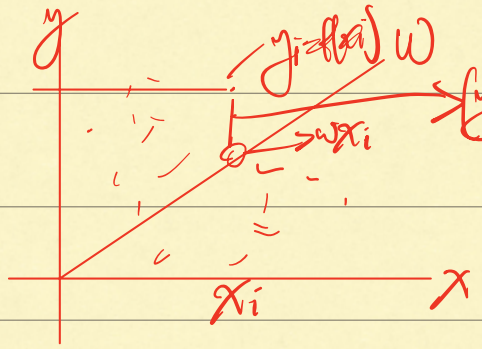
$$\min_{\boldsymbol{W},\boldsymbol{b}} \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{b}\|_2^2,$$

ii) (8 pts) Show how to use gradient descent to solve the problem. (Please state at least one possible Stopping Criterion)

iii) (8 pts) We further suppose that $x_1, x_2, \ldots, x_N$ are drawn from $\mathcal{N}(\mu, \sigma^2)$. Show that the maximum likelihood estimation (MLE) of σ^2 is $\hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{MLE})^2$.

i. $x_i \in \mathbb{R}^d$ $y_i \in \mathbb{R}^k$ $i = 1 \ldots n$ $W \in \mathbb{R}^{kd}$ $b \in \mathbb{R}^k$

Closed form solution: $(X^T X)^{-1} X^T Y$



$(y_i - wx_i)^2 =$ error

Here, in this question, $e = (y_i - (wx_i + b))^2$

$\begin{bmatrix} w \\ b \end{bmatrix}$ $[x, b]$ $wx + b$ $\begin{bmatrix} w \\ b \end{bmatrix}$ $[x, 1]$

$\mathbb{R}^{k \times (d+1)}$ $\mathbb{R}^{n \times (d+1)}$

minimize w, b so

thus, we let $\hat{w} = \begin{bmatrix} w \\ b \end{bmatrix}$ and $\hat{x} = [x, 1]$ where $\hat{w}$ is now $\in \mathbb{R}^{(d+1)k}$ and $\hat{x} \in \mathbb{R}^{(d+1)N}$

and $Y = \begin{bmatrix} y_1 \\ y_n \end{bmatrix}$. $\|y_i - wx_i - b\|_2^2 \to [Y - \hat{w}\hat{x}]^2 = [Y - \hat{w}\hat{x}]^T [Y - \hat{w}\hat{x}]$

$Y^T Y - Y^T \hat{w}\hat{x} - \hat{w}\hat{x}^T Y + (\hat{w}\hat{x})^T \hat{w}\hat{x}$

$\frac{d}{dw}\left(Y^T Y - 2\hat{w}\hat{x}^T Y + (\hat{w}\hat{x})^T \hat{w}\hat{x}\right) \to (\hat{x}^T \hat{w}^T)(\hat{w}\hat{x})$

$\hat{x}^T \hat{w}^T \hat{w} \hat{x}$

$-2\hat{x}^T Y + 2\hat{x}^T \hat{x}\hat{w} = 0$

$2\hat{x}^T \hat{x}\hat{w} = 2\hat{x}^T \hat{Y} \to \hat{x}^T \hat{x}\hat{w} = \hat{x}^T Y$

$\hat{w} = (\hat{x}^T \hat{x}^{-1}) \hat{x}^T Y = \begin{bmatrix} w^* \\ b^* \end{bmatrix}$

ii. Gradient Descent : $w_{k+1} = w_k - \alpha \frac{dL}{dw}$

$$= w_k - \alpha X^T(Xw - y)$$

$$= w_k - \alpha(-2X^T(Y - wX - b))$$

$$= w_k + 2\alpha X^T(Y - wX - b)$$

iii. $f(x_i|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i|\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\prod_{i=1}^{n} (\sigma^2)^{-1/2} \cdot \prod_{i=1}^{n} (2\pi)^{-1/2} \cdot \prod_{i=1}^{n} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \qquad a^n \cdot a^m = a^{n+m}$$

$$(\sigma^2)^{-n/2} \cdot (2\pi)^{-n/2} \cdot \left(\exp\left(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right)\right) \rightarrow \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Using log,

$$\log(L(\mu,\sigma^2)) = \log\left((\sigma^2)^{-n/2} \cdot (2\pi)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right)$$

$$l(\mu,\sigma^2) = \log((\sigma^2)^{-n/2}) + \log((2\pi)^{-n/2}) + \log\left(\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right)$$

$$= -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad -\frac{n}{2}(\sigma^2)^{-1} + \frac{1}{2}\cdot(\sigma^2)^{-2}$$

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - (\sigma^2)^{-3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(x_i) - n\mu\right)$$

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial \mu \partial \sigma^2} = -\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\left.\frac{\partial l}{\partial \mu}\right|_{(\mu,\sigma^2)=(\hat{\mu},\hat{\sigma}^2)} = 0$$

$$\cancel{\frac{1}{\hat{\sigma}^2}}\left(\sum_{i=1}^{n} x_i - n\hat{\mu}\right) = 0$$

$$\sum_{i=1}^{n} x_i - n\hat{\mu} = 0 \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(x_i)$$

$$\frac{\partial l}{\partial \sigma^2}\Big|_{(\mu,\sigma^2)=(\hat{\mu},\hat{\sigma}^2)} = 0$$

$$-\frac{n}{2}(\hat{\sigma}^2)^{-1} + \frac{1}{2}(\hat{\sigma}^2)^{-2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = 0$$

$$\left(\frac{n}{2}(\hat{\sigma}^2)^{-1} = \frac{1}{2}(\hat{\sigma}^2)^{-2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2\right)(\hat{\sigma}^2)^2$$

$$\frac{n}{2}(\hat{\sigma}^2) = \frac{1}{2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

1.2. (Support Vector Machine, 25 points) Given two positive samples $x_1 = (3,3)^T$, $x_2 = (4,3)^T$, and one negative sample $x_3 = (1,1)^T$, find the maximum-margin separating hyperplane and support vectors.

Solution steps:

i) Formulating the Optimization Problem (5 pts)

ii) Constructing the Lagrangian (5 pts)

iii) Using KKT Conditions (5 pts)

iv) Solving the Equations (5 pts)

v) Determining the Hyperplane Equation and Support Vectors (5 pts)

i. Hyperplane: $f_{w,b}(x) = w^T x + b = 0$

Maximizing margin: $\min_{w,b} \frac{1}{2}\|w\|^2$

s.t. $y_i(w^T x + b) \geq 1 \quad \forall i \in [1,2,3]$ → optimization problem

ii. $L(w,b,\lambda) = f(x) - \sum \lambda_i g_i(x) + \sum \eta_j h_j(x)$ $\quad g_i(x) \leq 0 \quad h_i(x) = 0 \quad 1 - y_i(w^T x + b) \leq 0$

$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{3} \lambda_i (1 - y_i(w^T x + b))$

iii. Main: $\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \lambda_i y_i x_i$

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

Dual: $\lambda_i \geq 0$ $\quad$ Primal: $1 - y_i(w^T x_i + b) \leq 0$

Complementary: $\lambda_i(1 - y_i(w^T x_i + b)) = 0 \quad \forall i,\ y_i \in (1,-1)$

iv. $x_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} y_1 = 1 \quad x_2 = \begin{bmatrix} 4 \\ 3 \end{bmatrix} y_2 = 1 \quad x_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} y_3 = -1$

$w = \sum_{i=1}^{3} \lambda_i y_i x_i = \lambda_1 \begin{bmatrix} 3 \\ 3 \end{bmatrix} + \lambda_2 \begin{bmatrix} 4 \\ 3 \end{bmatrix} - \lambda_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3\lambda_1 + 4\lambda_2 - \lambda_3 \\ 3\lambda_1 + 3\lambda_2 - \lambda_3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$\sum_{i=1}^{3} \lambda_i y_i = 0 \rightarrow \lambda_1 + \lambda_2 - \lambda_3 = 0 \rightarrow \boxed{\lambda_3 = \lambda_1 + \lambda_2}$

by complementary, $\lambda_1(1 - (w^T x_1 + b)) = 0$

$\lambda_2(1 - (w^T x_2 + b)) = 0$

$$\lambda_3(1+(w^T x_3+b))=0$$

Case 1, $\lambda_3=0$ $\rightarrow$ Assumption that λ_3 is not a support vector

then $\lambda_{1,2}\geq 0$

$\lambda_3=\lambda_1+\lambda_2 \rightarrow \lambda_1=-\lambda_2$ which violates $\lambda_{1,2}\geq 0$

Case 2, $\lambda_2=0$, $\lambda_{1,3}>0$

then $\lambda_3=\lambda_1$ by main and $\lambda_2=\lambda_3-\lambda_1$

From complementary, and using $\lambda_3=\lambda_1=\alpha$

$$w^T x_1+b=1 \rightarrow w\cdot x_1+b=[3\lambda_1-\lambda_3, 3\lambda_1-\lambda_3]\begin{pmatrix}3\\3\end{pmatrix}+b=1$$

$$12\alpha+b=1$$

$$w^T x_3+b=-1 \rightarrow w\cdot x_3+b=[3\lambda_1-\lambda_3, 3\lambda_1-\lambda_3]\begin{pmatrix}1\\1\end{pmatrix}+b=-1$$

$$4\alpha+b=-1$$

$$\begin{array}{r} 12\alpha+b=1 \\ -\ \underline{12\alpha+3b=-3} \\ -2b=4 \end{array} \quad b=-2$$

$12\alpha=3$ $\quad$ $4\alpha=1$

thus, $\alpha=\frac{3}{12}=\frac{1}{4}$ $\quad$ $\alpha=\frac{1}{4}$

$$w=\begin{pmatrix}w_1\\w_2\end{pmatrix}=\begin{pmatrix}3/4-1/4\\3/4-1/4\end{pmatrix}=\begin{pmatrix}1/2\\1/2\end{pmatrix}$$

v. Hyperplane: $f_{w,b}(x)=w^T x+b=0$

$$[1/2, 1/2]^T x+(-2)=0$$

Support Vectors: $x_i=\{i \mid \lambda_i>0\}$

thus, x_1 and x_3 are the support vectors

Programming

1.1

```python
regressor = LinearRegression()
regressor.fit(X_train, y_train)

w0 = regressor.intercept_
w = regressor.coef_

y_pred_test = regressor.predict(X_test)
y_pred_train = regressor.predict(X_train)

print(f"Bias (w0):", w0)
for i, coef in enumerate(w, start=1):
    print(f"w{i}: {coef}")

print("Predicted y values:")
for i, pred in enumerate(y_pred_test, start=1):
    print(f"y{i}: {pred}")
```

```
Bias (w0): 3.6136460266690658
w1: 0.015325690386247162
w2: 0.00025227640633962434
w3: 0.0007203868499651746
w4: 0.9991635645789794
w5: 0.9997402360310115
w6: 1.0006233993583622
w7: 0.9988323588429897
w8: 1.0000013015236562
w9: 1.000224548355873
w10: 0.9990396151482003
w11: 0.9993448460859436
Predicted y values:
y1: -56.111296879584195
y2: -173.51651970931798
y3: -6.770877912226837
y4: 209.5170904418719
y5: 116.8902978509852
y6: -100.2908452723578
y7: -310.12783900144865
y8: 501.3863019426078
y9: 244.11476780856407
y10: 18.56639325455041
```

1.2

```python
X_train_bias = np.c_[np.ones(X_train.shape[0]), X_train]
X_test_bias = np.c_[np.ones(X_test.shape[0]), X_test]

# w = (XTX)^(-1) * XTy
XTX = X_train_bias.T @ X_train_bias
XTy = X_train_bias.T @ y_train
XTX_inv = inv(XTX)
w = XTX_inv @ XTy

w0 = w[0]
weights = w[1:]

y_pred = X_test_bias @ w

print(f"Bias (w0): {w0}")
for i, wei in enumerate(weights, start=1):
    print(f"w{i}: {wei}")

# Print the predicted values for the test set
print("Predicted y values:")
for i, pred in enumerate(y_pred, start=1):
    print(f"y{i}: {pred}")
```

```
Bias (w0): 3.613646026668923
w1: 0.015325690386246506
w2: 0.000252276406340882
w3: 0.0007203868499652336
w4: 0.999163564578979
w5: 0.9997402360310119
w6: 1.0006233993583626
w7: 0.998832358842989
w8: 1.0000013015236553
w9: 1.000224548355873
w10: 0.999039615148201
w11: 0.9993448460859438
Predicted y values:
y1: -56.111296879584394
y2: -173.51651970931823
y3: -6.770877912227036
y4: 209.51709044187186
y5: 116.89029785098498
y6: -100.29084527235781
y7: -310.1278390014488
y8: 501.38630194260753
y9: 244.11476780856384
y10: 18.56639325455026
```

Technical Report ?

Q2.2.1 Split training set and test set:
Training set: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135]
Test set: [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150]

Q2.2.2 Calculation using Standard SVM Model:
total training error: 0.05714285714285714, total testing error: 0.08888888888888889

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: [-0.22149841131541403, 0.5197614640017161, -0.6204881374367106, -0.7018619843079239], b: -0.860042567927232
support vector indices: [58, 65, 24, 26]

class Iris-versicolor:

training error: 0.04761904761904762, testing error: 0.06666666666666667
w: [1.2403741375019308, -1.5570912008406594, -1.2096160736109596, -0.13571915771171916], b: -0.7952621398225709
support vector indices: [2, 4, 9, 10, 13, 14, 26, 31, 32, 35, 102, 103, 104, 106, 107, 108, 109, 112, 113, 114, 117, 119, 120, 123, 124, 126, 127, 129, 130, 131, 133, 134, 135, 51, 52, 53, 55, 56, 57, 58, 59, 60, 62, 64, 65, 66, 67, 68, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85]

class Iris-virginica:
training error: 0.009523809523809525, testing error: 0.0
w: [-0.887004583972157, -2.6477127808175283, 9.592181824380532, 10.405861852283124], b: -14.043043467030827
support vector indices: [71, 73, 78, 84, 127, 128, 134, 135]

Linear separable classes: ['Iris-setosa']

Q2.2.3 Calculation using SVM with Slack Variables (C = 0.25 × t, where t = 1,...,4):

C=0.25,
total training error: 0.08571428571428572, total testing error: 0.08888888888888889

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: [-0.2066 0.5757 -0.566 -0.595], b: -0.7129
support vector indices: [58, 65, 80, 9, 24, 26]
slack variables: [0.006352642943328601, 0.13320812782976255, 0.0, 0.0001248379583259629, 0.0, 0.0840452868089453]

class Iris-versicolor:
training error: 0.0761904761904762, testing error: 0.06666666666666667
w: [0.3367 -0.9205 -0.1189 -0.3382], b: -0.8506
support vector indices: [2, 3, 4, 9, 10, 13, 14, 26, 31, 35, 102, 103, 104, 106, 107, 108, 109, 112, 113, 114, 115, 117, 119, 120, 122, 123, 124, 127, 128, 129, 130, 131, 133, 134, 135, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85]
slack variables: [0.4778429972383941, 0.00014149461619816783, 0.14910568019877257, 0.4802708195444034, 0.3142625696430478, 0.48316881613391294, 0.3033220388562581, 0.5045060576033948, 0.22237386183705155, 0.3142625696430478, 0.42833983185735314, 0.19739482751075177, 0.23574216741159926, 0.3506521902440105, 0.6036045389885564, 0.5887883683308051, 1.1911451726559954, 0.6547917304082969, 0.10411093914357672, 0.7541791523074703, 0.0, 0.12008839583012043, 1.088967365875324, 1.7071203693584343, 0.11421875683050886, 0.833679511761723, 0.6867069638482045, 0.4511798893551555, 0.00014144709795360644, 0.3420658359426343, 0.47041788144756613, 0.7989739171994487, 0.29678224940992837, 0.6070463556821654, 0.9436109597439067, 1.8502429117482373, 2.1219783968318833, 1.746562312698373, 0.5378444198930782, 1.2798013408629618, 1.5022489552589753, 2.422731118765931, 0.7974254511573703, 1.3514930606523219, 1.5052205976690822, 1.897391853125622, 0.0, 1.6032128390070746, 1.6845972006632348, 1.7479574932595527, 2.0372072169454887, 1.0976343529873218, 0.17972732718669138, 0.805105446538294, 2.4775600555242425, 1.3091811165297236, 0.7730089555697676, 1.3104290054497094, 1.4114665350184041, 1.5856986004047866, 1.1279391586908882, 1.7214757141823018, 1.6751594859047727, 0.8954913383159604, 0.636199200047428, 0.5842682598786267, 1.1749068187805323, 1.355893873270156, 2.117122752219869]

class Iris-virginica:
training error: 0.009523809523809525, testing error: 0.0
w: [0.1078 -0.4612 1.3231 1.5083], b: -2.0673
support vector indices: [51, 52, 53, 54, 55, 57, 59, 64, 67, 69, 71, 73, 77, 78, 79, 84, 85, 102, 104, 107, 111, 112, 114, 117, 120, 122, 124, 126, 127, 128, 130, 132, 134, 135]
slack variables: [0.02325503202462209, 0.0004529764931029412, 0.46167260218241557,

0.023440508662464055, 0.492526045148717, 0.23622120223325438, 0.0001259253273899752, 0.21205214085850832, 0.10072284731734182, 0.9881048301829618, 0.7642059650782085, 0.9928360302402908, 0.47693579808270936, 1.01525223342572, 0.25323725632286465, 1.1016847153732705, 0.07512797891899115, 0.31810083030756076, 0.28881471652939616, 1.0783629492210998, 0.5332057110515986, 0.09336233387019899, 0.00028942480440541374, 0.43852146596101926, 0.6676053101096198, 0.39103782125119846, 0.6040039914598307, 0.18170404287866582, 0.7921030434889809, 0.9335728764846443, 0.530881471515011, 0.000289372441068636, 1.1631840951438648, 0.8181813319246451]

C=0.5,
total training error: 0.06666666666666667, total testing error: 0.044444444444444446

---

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: [-0.2215 0.5198 -0.6205 -0.7019], b: -0.86
support vector indices: [58, 65, 24, 26]
slack variables: [0.0, 0.0006537330018741017, 0.0003220104153144332, 9.62118586322891e-06]

---

class Iris-versicolor:
training error: 0.05714285714285715, testing error: 0.022222222222222223
w: [0.4454 -1.0018 -0.2309 -0.3336], b: -0.8228
support vector indices: [2, 3, 4, 9, 10, 13, 14, 26, 31, 35, 102, 103, 104, 106, 107, 108, 109, 112, 113, 114, 117, 119, 120, 122, 123, 124, 126, 127, 128, 129, 130, 131, 133, 134, 135, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85]
slack variables: [0.5327738680235097, 0.0, 0.14121581544378792, 0.4885997425240969, 0.34443014882031164, 0.5245776659278008, 0.2990478179914533, 0.5598042876222735, 0.23400997604525076, 0.34443014882031164, 0.4318444002466659, 0.2660597883868504, 0.23603037800425275, 0.43994520778456325, 0.5631282894685192, 0.6741789667422663, 1.3019748745707593, 0.723137989185719, 0.15914631532128376, 0.7874245294226954, 0.13455770903413744, 1.2451078320675366, 1.8295317214112767, 0.08721587116661689, 0.9647212539720992, 0.7665862156286205, 0.0, 0.506555645054325, 0.0006138163698210519, 0.37500276039617686, 0.5551056882512858, 0.9282864269481942, 0.3303463286238284, 0.6546442528354502, 0.9692179688801776, 1.7594426251407338, 2.095392645852129, 1.6626848612095548, 0.4363965568786847, 1.175094943755148, 1.495692044020295, 2.4388129696027354, 0.7492543586683234, 1.2530184901832986, 1.5070559584841206, 1.8807444061066816, 1.5748491984305995, 1.652433757517355, 1.6591781615000567, 2.0780356291148303, 1.0371366424556956, 0.00020747517932528936, 0.721498040013064, 2.5323584319134285, 1.2197259728747802, 0.6592664013725602, 1.2654472910450234, 1.3199904365154516, 1.4919417515272249, 0.9977028246479412, 1.6505250557841264, 1.646536049801696, 0.7924335896749288, 0.5413505229055786, 0.4837829839985792, 1.1006272917310715, 1.328481036700184, 2.183740896850952]

---

class Iris-virginica:
training error: 0.009523809523809525, testing error: 0.0
w: [0.0341 -0.6345 1.756   1.8244], b: -2.5081
support vector indices: [53, 54, 55, 57, 64, 67, 69, 71, 73, 77, 78, 79, 84, 85, 102, 104, 107, 111, 117, 120, 122, 124, 126, 127, 128, 130, 134, 135]
slack variables: [0.3139594709815823, 0.000340503251962776, 0.4214139938459833, 0.058783831923995056, 0.11975917800988767, 0.008023075809602709, 1.1474434499168997, 0.7686547276863274, 1.1260211676186893, 0.3856625988134059, 1.0319262874950312, 0.16360649379358128, 1.275693896459234, 0.0, 0.0, 0.011569727823290687, 0.834859725320634, 0.42403410589176893, 0.24104150524678358, 0.3697718623698978, 0.09925168673055484, 0.42001942670986114, 0.0006083271170362892, 0.6616350852600266, 0.8462900769062038, 0.4066724566292388, 1.0958013184408135, 0.5784809973091036]

C=0.75,
total training error: 0.06666666666666667, total testing error: 0.044444444444444446

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: [-0.2215 0.5198 -0.6205 -0.7019], b: -0.86
support vector indices: [58, 65, 24, 26]
slack variables: [0.0, 0.0006537330018741017, 0.0003220104153144332, 9.62118586322891e-06]

class Iris-versicolor:
training error: 0.05714285714285715, testing error: 0.022222222222222223
w: [0.5153 -1.0113 -0.2977 -0.3483], b: -0.8342
support vector indices: [2, 3, 4, 9, 10, 13, 14, 26, 31, 35, 102, 103, 104, 106, 107, 108, 109, 112, 113, 114, 117, 119, 120, 122, 123, 124, 126, 127, 129, 130, 131, 133, 134, 135, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85]
slack variables: [0.5499373768523352, 0.0, 0.12767497283672735, 0.46639731067395973, 0.35774527831730285, 0.5354299724234065, 0.2796489861614009, 0.5777877035797907, 0.23331677741651324, 0.35774527831730285, 0.3580131820769149, 0.259894076560655, 0.18277294822041967, 0.4490868613228052, 0.44523382554424706, 0.6776878703694487, 1.2827848422140473, 0.6915856461912453, 0.1430489956283768, 0.7112429243430745, 0.09952678234161683, 1.2557443764124394, 1.7944075356430367, 0.00019933361511925707, 0.9845854038907128, 0.7436664432621671, 0.0, 0.4769869949512543, 0.32618837776093823, 0.5708706696552686, 0.9476714467568185, 0.27955135471248505, 0.6281013404491046, 0.9135783990850378, 1.719390727495009, 2.099600214657806, 1.6382842075622555, 0.4737521231058378, 1.1663745527213334, 1.5456088760684632, 2.462857837142078, 0.8063603374085103, 1.2341381503555593, 1.5759036630455157, 1.91101705850277, 1.6031443611658598, 1.6791119213797008, 1.6307007875937498, 2.1443914920597864, 1.0557831067373606, 6.835752742284651e-05, 0.7470507415736146, 2.595174572314939, 1.2177959142841526, 0.6720583451754157, 1.2876882438604293, 1.3064858541854243, 1.4696631438626075, 0.9695983479906654, 1.6483120877803774, 1.6776317109417667, 0.7948631607564528, 0.569366047467519, 0.506081974044099, 1.1157630520843522, 1.3797868938229567, 2.2666803470145367]

class Iris-virginica:
training error: 0.009523809523809525, testing error: 0.0
w: [-0.0908 -0.5879 1.9794 1.9729], b: -2.7242
support vector indices: [53, 55, 57, 64, 69, 71, 73, 77, 78, 79, 84, 85, 107, 111, 117, 120, 122, 124, 126, 127, 128, 130, 134, 135]
slack variables: [0.15153575355253346, 0.25012848268914256, 0.000676906497690144, 0.010630207414454418, 0.9467710941502929, 0.8118874469580772, 0.9911903273942331, 0.17500313647546095, 0.9411860281414364, 0.06419528775665295, 1.250677335825471, 0.0, 0.7723179166924723, 0.3925225108301773, 0.21976781356733932, 0.4783108221115926, 0.0001952574036527821, 0.4747500410221899, 0.00019551832664399527, 0.703803767697651, 0.8406766632643308, 0.49147020250376805, 1.1749534335727398, 0.6058388323119823]

C=1.0,
total training error: 0.06666666666666667, total testing error: 0.044444444444444446

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: [-0.2215 0.5198 -0.6205 -0.7019], b: -0.86
support vector indices: [58, 65, 24, 26]
slack variables: [0.0, 0.0006537330018741017, 0.0003220104153144332, 9.62118586322891e-06]

class Iris-versicolor:
training error: 0.05714285714285715, testing error: 0.022222222222222223
w: [0.5335 -1.0093 -0.3441 -0.3167], b: -0.8324
support vector indices: [2, 3, 4, 9, 10, 13, 14, 26, 31, 35, 102, 103, 104, 106, 107, 108, 109, 112, 113, 114, 117, 119, 120, 122, 123, 124, 126, 127, 129, 130, 131, 133, 134, 135, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85]
slack variables: [0.5505681259143156, 0.0, 0.11970447596851852, 0.45578695080586706,

0.35199997385783643, 0.5296604255458844, 0.2708717133399171, 0.5753920684012769, 0.22706764655863665, 0.35199997385783643, 0.352939247869948, 0.2719875473644555, 0.17219855676691764, 0.4538225454154319, 0.42689446684195076, 0.6705375526400843, 1.2738161888021915, 0.6942671881791033, 0.15903886282610136, 0.7099294230757347, 0.09631576172079603, 1.2615131518266254, 1.7769657858903192, 0.0006985839087955137, 0.9837236610646263, 0.7503132019656309, 0.0, 0.484528008226192, 0.33004288145711835, 0.5664922714476581, 0.9516449327997913, 0.28764978226110793, 0.6172928197624019, 0.8803378225632903, 1.7071460607905407, 2.090867100295691, 1.629593993914726, 0.4840905508661759, 1.1599134384149794, 1.5622952079997665, 2.4567606788473717, 0.8237698736073455, 1.233549263172301, 1.584038911071754, 1.9062101670741556, 1.611684788729247, 1.6741729734292528, 1.6176103511538704, 2.1538278513837135, 1.0731385739785915, 0.0002404722870690268, 0.7602107697984861, 2.5922656813221328, 1.212899867048284, 0.6790618550188541, 1.3051757676234996, 1.3024355766849383, 1.4591883280045619, 0.9660947976031637, 1.6384973972606058, 1.6789018304122032, 0.7992859795968053, 0.5825503181104297, 0.5209187903756811, 1.1194479152931083, 1.393279855413176, 2.280429450512596]

class Iris-virginica:
training error: 0.009523809523809525, testing error: 0.0
w: [-0.2103 -0.6561 2.2408 2.0126], b: -2.8706
support vector indices: [53, 55, 64, 67, 69, 71, 73, 77, 78, 79, 84, 85, 107, 111, 117, 120, 124, 126, 127, 128, 130, 132, 134, 135]
slack variables: [0.03230731643235485, 0.18870023948449433, 0.00022121446362133312, 0.0, 1.0031114044230325, 0.820812179195265, 1.046871199764618, 0.09494186318882303, 0.8905433684133279, 0.044065763445851225, 1.3534833519572476, 0.049670178487499106, 0.5659567383903044, 0.4141801901290245, 0.1636813269827324, 0.3205648811456774, 0.43304816351474074, 0.00024396925922509638, 0.6775102941802529, 0.8155455368972824, 0.5014369706543547, 0.0, 1.1349760460842235, 0.4398297880282689]

Q2.2.4 Calculation using SVM with Kernel Functions:

(a) 2nd-order Polynomial Kernel
total training error: 0.24761904761904763, total testing error: 0.26666666666666666

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: , b: -2.338
support vector indices: [101, 103, 104, 105, 106, 108, 109, 110, 112, 113, 114, 115, 116, 119, 121, 123, 126, 129, 131, 133, 2, 3, 4, 6, 8, 9, 10, 12, 13, 14, 19, 21, 24, 25, 26, 27, 29, 30, 31, 32, 35]

class Iris-versicolor:
training error: 0.0, testing error: 0.0
w: , b: 1.2868
support vector indices: [102, 104, 107, 111, 112, 114, 117, 120, 122, 124, 125, 126, 127, 128, 130, 134, 135, 51, 52, 53, 54, 55, 57, 58, 59, 64, 66, 69, 71, 73, 76, 77, 78, 79, 84]

class Iris-virginica:
training error: 0.24761904761904766, testing error: 0.24444444444444446
w: , b: -1.1291
support vector indices: [1, 2, 3, 4, 7, 8, 9, 10, 12, 13, 14, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 35, 51, 52, 53, 54, 57, 58, 60, 61, 66, 71, 78, 81, 82, 85, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135]

(b) 3rd-order Polynomial Kernel
total training error: 0.10476190476190476, total testing error: 0.06666666666666667

class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: , b: -0.9896
support vector indices: [58, 65, 68, 85, 21, 24, 26, 32]

class Iris-versicolor:
training error: 0.09523809523809526, testing error: 0.044444444444444446
w: , b: -0.9449
support vector indices: [4, 7, 9, 12, 14, 21, 24, 25, 26, 27, 28, 29, 32, 102, 103, 104, 105, 106, 107, 108, 109, 111, 112, 113, 114, 117, 119, 120, 121, 124, 127, 128, 129, 131, 133, 134, 135, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85]

class Iris-virginica:
training error: 0.009523809523809525, testing error: 0.0
w: , b: -1.0945
support vector indices: [51, 52, 53, 54, 55, 57, 59, 60, 64, 66, 69, 71, 73, 76, 77, 78, 79, 84, 102, 104, 107, 111, 112, 114, 116, 117, 120, 122, 124, 125, 126, 127, 128, 130, 134, 135]

(c) Radial Basis Function Kernel with σ = 1
total training error: 0.02857142857142857, total testing error: 0.022222222222222223
class Iris-setosa:
training error: 0.0, testing error: 0.0
w: , b: -0.3511
support vector indices: [51, 57, 58, 61, 63, 65, 69, 85, 101, 107, 109, 110, 115, 118, 119, 132, 9, 15, 16, 19, 23, 26, 32, 33]
class Iris-versicolor:
training error: 0.009523809523809525, testing error: 0.0
w: , b: -0.399
support vector indices: [9, 16, 19, 23, 24, 26, 33, 102, 104, 107, 109, 111, 117, 119, 120, 122, 124, 126, 127, 128, 130, 132, 134, 135, 51, 53, 54, 55, 57, 58, 60, 61, 63, 69, 71, 73, 77, 78, 79, 84, 85]
class Iris-virginica:
training error: 0.01904761904761905, testing error: 0.022222222222222223
w: , b: -0.2968
support vector indices: [9, 14, 15, 16, 19, 21, 23, 33, 51, 53, 54, 55, 57, 60, 61, 64, 69, 71, 73, 77, 78, 79, 84, 85, 101, 102, 107, 109, 110, 111, 114, 117, 119, 120, 122, 124, 126, 127, 128, 130, 132, 134, 135]

(d) Sigmoidal Kernel with σ = 1
total training error: 0.37142857142857144, total testing error: 0.26666666666666666
class Iris-setosa:
training error: 0.0, testing error: 0.022222222222222223
w: , b: -0.583584541469575
support vector indices: [58, 61, 65, 80, 81, 82, 132, 2, 4, 9, 13, 14, 26]
class Iris-versicolor:
training error: 0.20952380952380956, testing error: 0.08888888888888889
w: , b: -0.15979735636883713
support vector indices: [16, 19, 24, 101, 103, 105, 106, 107, 108, 109, 110, 112, 113, 117, 119, 120, 121, 123, 125, 126, 129, 130, 131, 133, 51, 52, 53, 55, 56, 57, 59, 62, 64, 66, 67, 69, 71, 72, 73, 74, 75, 76, 77, 78, 79, 84, 85]
class Iris-virginica:
training error: 0.1619047619047619, testing error: 0.15555555555555556
w: , b: -3.538003779016435
support vector indices: [51, 52, 53, 55, 56, 57, 59, 62, 64, 66, 67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 84, 85, 101, 103, 105, 106, 107, 108, 109, 110, 112, 113, 118, 119, 120, 121, 123, 125, 126, 129, 130, 131, 132, 133]

Notes

w is orthogonal to the hyperplane

$w^T x_1 + b = 0$, $w^T x_2 + b = 0$ thus $w^T(x_1 - x_2) = 0$

Hyperplane: $f_{w,b}(x) = w^T x + b = 0$ where $w = [w_1, \dots w_n]^T$, $b = w_0$, $x = [x_1, \dots, x_n]^T$

0 has distance $\frac{-b}{\|w\|}$ to a hyperplane $w^T x + b = 0$

Project x_i to the unit direction $\frac{w}{\|w\|}$, $\left(\frac{w}{\|w\|}\right)^T x_i = \frac{-b}{\|w\|}$ since $w^T x_i + b = 0$

$\text{Proj} = x_i^T \frac{w}{\|w\|} = \frac{w^T x_i}{\|w\|} = -\frac{b}{\|w\|}$, $w^T x + b = 0 \rightarrow w^T x = -b$

x has distance $\frac{|f_{w,b}(x)|}{\|w\|}$ to $f_{w,b}(x) = w^T x + b = 0$

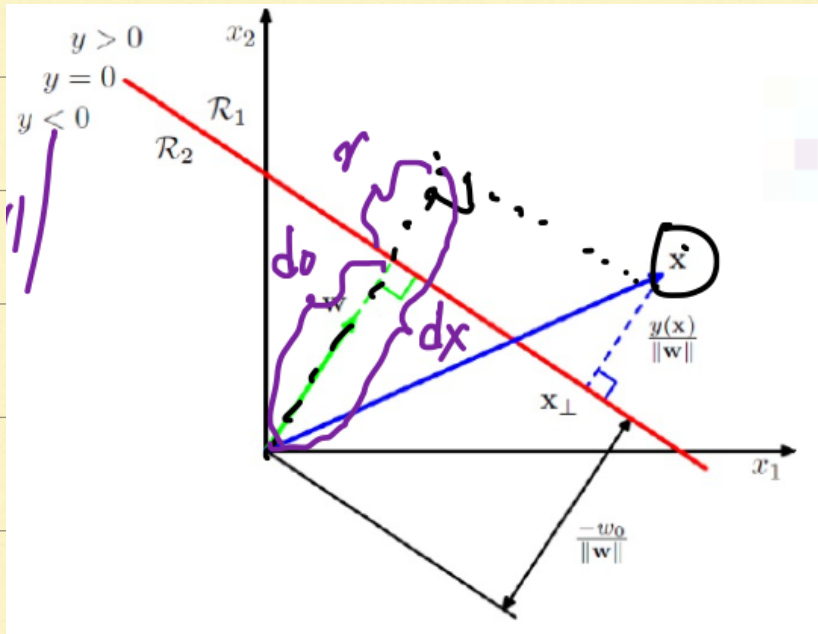
let $x = x_{proj} + r \frac{w}{\|w\|}$ where $|r|$ is the distance

$w^T x + b = w^T x_{proj} + r \cdot \frac{w^T w}{\|w\|} + b$

$f_{w,b}(x) = 0 + r\|w\| \rightarrow |r| = \frac{|f_{w,b}(x)|}{\|w\|}$

$r = |d_x - d_0| = \left| x^T \left(\frac{w}{\|w\|}\right) - \frac{-b}{\|w\|} \right| = \frac{|w^T x + b|}{\|w\|} = \frac{|f_{w,b}(x)|}{\|w\|}$

$f_{w,b}(x) = w^T x + w_0$, $b = w_0$



$$\min_{\bar{x}} \frac{1}{2}\|x-\bar{x}\|^2 \qquad \mathcal{L} = \frac{1}{2}\|x-\bar{x}\|^2 + \lambda(w^T x + b)$$

$$\text{s.t. } w^T\bar{x} + b = 0 \qquad \frac{\partial \mathcal{L}}{\partial \bar{x}} = -(x-\bar{x}) + \lambda w = 0 \rightarrow \lambda w = x - \bar{x}$$

$$\lambda w^T w = w^T x - w^T \bar{x} = w^T x + b - (w^T\bar{x} + b) = w^T x + b$$

$$\lambda \|w\|^2 = w^T x + b$$

$$\|x-\bar{x}\| = \lambda\|w\| = \frac{w^T x + b}{\|w\|} = \frac{|f_{w,b}(x)|}{\|w\|}$$

$$\gamma^i = \min_i \frac{|f_{w,b}(x_i)|}{\|w\|} \rightarrow \gamma = \min_i \frac{y_i f_{w,b}(x_i)}{\|w\|}$$ if $f_{w,b}$ incorrect for some x_i, margin < 0

$$\max_{w,b} \gamma = \max_{w,b} \min_i \frac{y_i f_{w,b}(x_i)}{\|w\|} = \max_{w,b} \min_i \frac{y_i(w^T x_i + b)}{\|w\|}$$

if $y_{i^*}(w^T x_{i^*} + b) = 1$ where x_{i^*} = closest to the hyperplane,

then $y_i(w^T x_i + b) \geq 1 \rightarrow$ margin $= \frac{1}{\|w\|}$

$$\min_{w,b} \frac{1}{2}\|w\|^2$$
$$\text{s.t. } y_i(w^T x + b) \geq 1 \ \forall i$$

$\rightarrow$ Solving this, we obtain w^*, b^*

Given w^*, b^*, for x_t (new test data),

we predict it as $+1$ if $(w^*)^T x_t + b^* > 0$, -1 otherwise

Hinge loss: $\max(0, 1 - y_i(w^T x + b))$

Hinge loss isn't smooth, therefore objective function of SVM becomes:

$$\min_{w,b} \frac{1}{2}\sum_{i=1}^{n} w_i^2$$
$$\text{s.t. } w^T x_i + b \geq 1,\ y_i = 1$$
$$w^T x_i + b < 1,\ y_i = -1$$

$\rightarrow$

$$\min_{w,b} \frac{1}{2}\|w\|^2$$
$$\text{s.t. } y_i(w^T x + b) \geq 1 \ \forall i$$

$p_i = \frac{w^T x_i + b}{\|w\|}$: Projection length of x_i on w or distance from x_i to $w^T x + b = 0$

$$w^T x_i + b = p_i \cdot \|w\|$$

Of of SVM: $\min_{w,b} \frac{1}{2}\|w\|^2$

$$\text{s.t. } y_i \cdot p_i \cdot \|w\| \geq 1,\ \forall i$$

the larger p_i, the smaller $\|w\|$

thus, we prefer large margin

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t. } y_i(w^T x + b) \geq 1 \ \forall i \quad \rightarrow \quad \min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t. } 1 - y_i(w^T x + b) \leq 0 \ \forall i$$

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2}\|w\|^2 + \sum_i^n \lambda_i (1 - y_i(w^T x + b))$$

Main: $\frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_i^n \lambda_i y_i x_i$

$\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_i^n \lambda_i y_i = 0$

Dual: $\lambda_i \geq 0$ Primal: $1 - y_i(w^T x_i + b) \leq 0 \ \forall i$

Complementary: $\lambda_i(1 - y_i(w^T x_i + b)) = 0 \ \forall i$

Thus, $\mathcal{L}(w,b,\lambda) = \frac{1}{2}\|w\|^2 + \sum_i^n \lambda_i - \sum_i^n \lambda_i y_i \left(\sum_j^n \lambda_j y_j x_j\right)^T x_i - \sum_i^n \lambda_i y_i b$

$$= \frac{1}{2}\|w\|^2 + \sum_i^n \lambda_i - \sum_{ij}^n \lambda_i \lambda_j y_i y_j x_i^T x_j - b \sum_i^n \lambda_i y_i$$

$$= \sum_i^n \lambda_i - \frac{1}{2}\|w\|^2$$

$$= \sum_i^n \lambda_i - \frac{1}{2} \sum_{ij}^n \lambda_i \lambda_j y_i y_j x_i^T x_j$$

Dual problem: $\max_{\lambda} \sum_i^n \lambda_i - \frac{1}{2} \sum_{ij}^n \lambda_i \lambda_j y_i y_j x_i^T x_j$

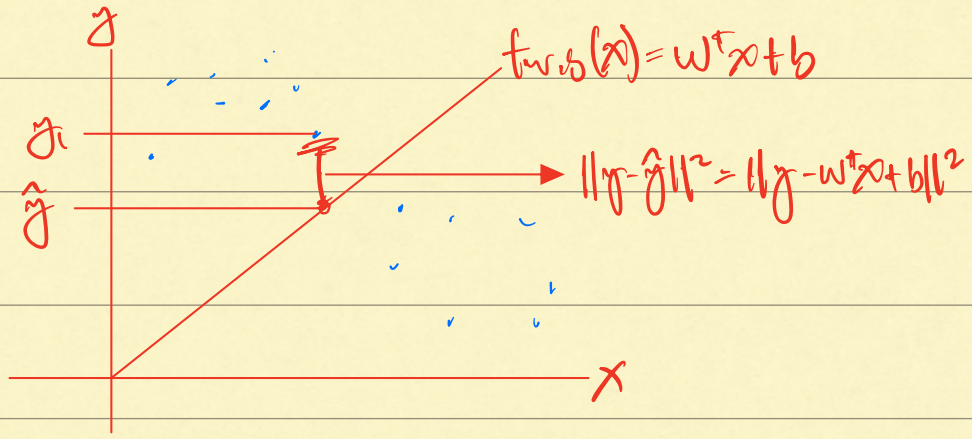
s.t. $\sum_i^n \lambda_i y_i = 0, \ \lambda_i \geq 0 \ \forall i$

replacing λ^* back into Main, we have $w^* = \sum_i^n \lambda_i^* y_i x_i$

Data points with $\lambda_i > 0$ is called support vectors

which locate at the hyperplanes $y_i(w^T x + b) = 1$

Support vectors set: $S = \{i \mid \lambda_i > 0\} \rightarrow$ Support Vector Machine


y
y_i
$\hat{y}$
x
$f_{w,b}(x) = w^T x + b$
$\|y - \hat{y}\|^2 = \|y - w^T x + b\|^2$

1.2. (Support Vector Machine, 25 points) Given two positive samples $x_1 = (3,3)^T$, $x_2 = (4,3)^T$, and one negative sample $x_3 = (1,1)^T$, find the maximum-margin separating hyperplane and support vectors.

Solution steps:

i) Formulating the Optimization Problem (5 pts)

ii) Constructing the Lagrangian (5 pts)

iii) Using KKT Conditions (5 pts)

iv) Solving the Equations (5 pts)

v) Determining the Hyperplane Equation and Support Vectors (5 pts)

i. Separating hyperplane equation: $wx+b=0$

Opti problem: $\min_{w,b} \frac{1}{2}\|w\|^2$

s.t. $wx_i+b \geq 1$

$wx_i+b \leq -1$ $\rightarrow$ s.t. $y_i(wx_i+b) \geq 1 \ \forall i, y_i \in (-1,1)$

s.t. $3w_1+3w_2+b \geq 1$

$4w_1+3w_2+b \geq 1$ $\rightarrow$ $-(3w_1+3w_2+b) \leq -1$, $-(\text{—}) \leq -1$

$w_1+w_2+b \leq -1$

ii. $L = f(x) + \sum \lambda_i g_i(x) + \sum \eta_j h_j(x)$ $\quad g_i(x) \leq 0$, $h_j(x) = 0$

$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{3} \lambda_i (y_i(wx+b)-1)$

$= \frac{1}{2}(w_1^2+w_2^2) - \lambda_1(3w_1+3w_2+b-1) - \lambda_2(4w_1+3w_2+b-1) + \lambda_3(w_1+w_2+b+1)$

iii. Main: $\frac{\partial L}{\partial w_1} = w_1 - 3\lambda_1 - 4\lambda_2 + \lambda_3 = 0 \rightarrow w_1 = 3\lambda_1 + 4\lambda_2 - \lambda_3$

$\frac{\partial L}{\partial w_2} = w_2 - 3\lambda_1 - 3\lambda_2 + \lambda_3 = 0 \rightarrow w_2 = 3\lambda_1 + 3\lambda_2 - \lambda_3$

$\frac{\partial L}{\partial b} = -\lambda_1 - \lambda_2 + \lambda_3 = 0 \rightarrow \lambda_3 = \lambda_1 + \lambda_2$

Dual: $\lambda_i \geq 0$

Complementary: $\lambda_i(y_i(wx+b)-1) = 0 \ \forall i, y \in (1,-1)$

Primal: $y_i(wx+b) \geq 1 \ \forall i, \in (1,-1)$

iv. $w_1 = 3\lambda_1 + 4\lambda_2 - \lambda_1 - \lambda_2 = 2\lambda_1 + 3\lambda_2$

$w_2 = 3\lambda_1 + 3\lambda_2 - \lambda_1 - \lambda_2 = 2\lambda_1 + 2\lambda_2$

$x_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ $x_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ $x_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$y_i = 1$ $\quad 1$ $\quad -1$

By complementary, $\lambda_i(y_i(wx_i + b) - 1) = 0 \quad \forall i$

Case 1, $\lambda_3 = 0$ → Assumption that λ_3 is not a support vector

$\lambda_{1,2} > 0$

From main, $\lambda_3 = \lambda_1 + \lambda_2 = 0$

$\lambda_1 = -\lambda_2$ which violates our case where $\lambda_{1,2} > 0$

Since theres no solution, our assumption about λ_3 is incorrect

Case 2, $\lambda_2 = 0$ $\quad \lambda_2 \neq$ support vector

$\lambda_{1,3} > 0$

From main, $\lambda_3 = \lambda_1$

By complementary, $\lambda_1(y_1(w\begin{bmatrix} 3 \\ 3 \end{bmatrix} + b) - 1) = 0$ $\quad \lambda_3(\overset{-1}{y_3}(w\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b) - 1) = 0$

$w\begin{bmatrix} 3 \\ 3 \end{bmatrix} + b = 1$ $\quad w\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = -1$

if $\lambda_2 = 0$, $w_1 = 3\lambda_1 - \lambda_3$ from main

$w_2 = 3\lambda_1 - \lambda_3$

$[3\lambda_1 - \lambda_3, 3\lambda_1 - \lambda_3]\begin{bmatrix} 3 \\ 3 \end{bmatrix} + b = 1$ $\quad [3\lambda_1 - \lambda_3, 3\lambda_1 - \lambda_3]\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = -1$

$\lambda_3 = \lambda_1$, $[2\lambda_3, 2\lambda_3]\begin{bmatrix} 3 \\ 3 \end{bmatrix} + b = 1$ $\quad [2\lambda_3, 2\lambda_3]\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = -1$

$12\lambda_3 + b = 1$ $\quad 4\lambda_3 + b = -1$

$12\lambda_3 + b = 1$
$-12\lambda_3 + 3b = -3$
$-2b = 4$
$b = -2$

Thus, $12\lambda_3 = 3$ $\quad 4\lambda_3 = 1$

$\lambda_3 = \frac{1}{4}$ $\quad \lambda_3 = \frac{1}{4}$