Solution Set: Support Vector Classifier

1. Our convex optimization problem takes the form:

$$\underset{(\beta_0,\beta,\varepsilon)\in\mathbb{R}^7}{minimize} \quad f(\beta_0,\beta_1,\beta_2,\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4) \qquad \text{given the constraint}$$

$g_i(\beta_0,\beta,\varepsilon)\le 0$ for $i = 1, 2, 3, 4$

and $h_i(\beta_0,\beta,\varepsilon)\le 0$ for $i = 1, 2, 3, 4$

where $(\beta_0,\beta,\varepsilon) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{4}\varepsilon_i$,

$g_i(\beta_0,\beta,\varepsilon) = 1-\varepsilon_i - y_i(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})$ for $= 1, 2, 3, 4$,
and $h_i(\beta_0,\beta,\varepsilon) = -\varepsilon_i$ for $i = 1, 2, 3, 4$

So
$$\begin{aligned}
g_1 &= 1-\varepsilon_1+(\beta_0)\\
g_2 &= 1-\varepsilon_2+(\beta_0+\beta_2)\\
g_3 &= 1-\varepsilon_3-(\beta_0-\beta_1)\\
g_4 &= 1-\varepsilon_4-(\beta_0+\beta_1)\\
h_1 &= -\varepsilon_1\\
h_2 &= -\varepsilon_2\\
h_3 &= -\varepsilon_3\\
h_4 &= -\varepsilon_4.
\end{aligned}$$

The dual Lagrangian is given by $L_D(\alpha) = \sum_{i=1}^{4}\alpha_i - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4}\alpha_i\alpha_j y_i y_j x_i^T x_j$.

So $L_D(x,\alpha) = (\alpha_1+\alpha_2+\alpha_3+\alpha_4) - \frac{1}{2}[\alpha_2^2+\alpha_3^2+\alpha_4^2-2\alpha_3\alpha_4]$

We want to maximize $L_D(\alpha)$ subject to the constraints $0\le\alpha_i\le C\ \forall\, i$ and $\alpha_1 y_1+\alpha_2 y_2+\alpha_3 y_3+\alpha_4 y_4 = 0$. That is, we need $0\le\alpha_i\le C\ \forall\, i$ and $-\alpha_1-\alpha_2+\alpha_3+\alpha_4 = 0$. These constraints give us a four-dimensional plane in the positive box $0\le\alpha_i\le C\ \forall\, i$.

Let $H = \{(\alpha_1,\dots,\alpha_4)\in\mathbb{R}^4 | 0\le\alpha_i\le C\ \forall\, i\ and\ -\alpha_1-\alpha_2+\alpha_3+\alpha_4 = 0\}$. We want to maximize $L_D(\alpha_1,\dots,\alpha_4)$ on H.

To find the maximum of $L_D(\alpha_1,\dots,\alpha_4)$ on H, we can use any computational software.

It turns out that, for $C = 2$, the maximum value of L_D on H is 6 and it occurs at $(\alpha_1,\dots,\alpha_4) = (2, 2, 2, 2)$.

$$\beta = \sum_{i=1}^{4}\alpha_i y_i x_i$$

$$\Rightarrow \qquad \beta = \begin{bmatrix}0\\-2\end{bmatrix}.$$

By complementary slackness, we have $\alpha_i\big(1-\varepsilon_i-y_i(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})\big) = 0\ \forall\, i = 1,\dots,4$

This gives us a system of 4 equations and 5 unknowns $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \beta_0$. Solving this system gives $\varepsilon_1 = 2, \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 0$, and $\beta_0 = 1$.

The equation of our hyperplane is given by $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$.

So we get $1 - 2X_2 = 0$

$$\Longrightarrow \qquad X_2 = \frac{1}{2}$$

Since $\alpha_i > 0$ for $i = 1, \dots, 4$, we have that each x_i satisfies $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = 1 - \varepsilon_i$. Hence, x_1, x_2, x_3, x_4 are all support vectors.

b) For $C = 4$, L_D has an absolute max value of 10 and it occurs at $(\alpha_1, \dots, \alpha_4) = (4, 2, 3, 3)$.

$\beta = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \beta_0 = 1, \varepsilon_1 = 2, \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 0$, and the hyperplane is $X_2 = \frac{1}{2}$, the same result we got for $C = 2$.

c) For $C = 1$, L_D has an absolute max value of $\frac{7}{2}$ and it occurs at $(\alpha_1, \dots, \alpha_4) = (1, 1, 1, 1)$.

$\beta = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$. The complementary slackness equations

$$\alpha_i\big(1 - \varepsilon_i - y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})\big) = 0 \ \forall \ i = 1, \dots, 4$$

give us a system of 4 equations and 5 unknowns $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \beta_0$. This system has more than one solution. One solution is $\beta_0 = 0, \varepsilon_1 = 1, \varepsilon_2 = 0, \varepsilon_3 = 1, \varepsilon_4 = 1$. The hyperplane is $X_2 = 0$. Another solution is $\beta_0 = 1, \varepsilon_1 = 2, \varepsilon_2 = 1, \varepsilon_3 = 0, \varepsilon_4 = 0$. The hyperplane is $X_2 = 1$.

2. Our convex optimization problem takes the form:

$\underset{(\beta_0,\beta,\varepsilon)\in\mathbb{R}^8}{minimize} \quad f(\beta_0,\beta_1,\beta_2,\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4,\varepsilon_5)$ given the constraint

$g_i(\beta_0,\beta,\varepsilon) \le 0$ for $i = 1, 2, 3, 4, 5$

and $h_i(\beta_0,\beta,\varepsilon) \le 0$ for $i = 1, 2, 3, 4, 5$

where $(\beta_0,\beta,\varepsilon) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{5}\varepsilon_i$,

$g_i(\beta_0,\beta,\varepsilon) = 1 - \varepsilon_i - y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$ for $= 1, 2, 3, 4$, 5,
and $h_i(\beta_0,\beta,\varepsilon) = -\varepsilon_i$ for $i = 1, 2, 3, 4, 5$

So

$$
\begin{aligned}
g_1 &= 1 - \varepsilon_1 - (\beta_0 + \beta_2)\\
g_2 &= 1 - \varepsilon_2 - (\beta_0 - \beta_2)\\
g_3 &= 1 - \varepsilon_3 + (\beta_0)\\
g_4 &= 1 - \varepsilon_4 + (\beta_0 + \beta_1 + \beta_2)\\
g_5 &= 1 - \varepsilon_5 + (\beta_0 + \beta_1 - \beta_2)\\
h_1 &= -\varepsilon_1\\
h_2 &= -\varepsilon_2\\
h_3 &= -\varepsilon_3\\
h_4 &= -\varepsilon_4\\
h_5 &= -\varepsilon_5
\end{aligned}
$$

The dual Lagrangian is given by $L_D(\alpha) = \sum_{i=1}^{5}\alpha_i - \frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5}\alpha_i\alpha_j y_i y_j x_i^T x_j$.

So $L_D(x,\alpha) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5) - \frac{1}{2}[\alpha_1^2 + \alpha_2^2 + 2\alpha_4^2 + 2\alpha_5^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_4 + 2\alpha_1\alpha_5 + 2\alpha_2\alpha_4 - 2\alpha_2\alpha_5]$

We want to maximize $L_D(\alpha)$ subject to the constraints $0 \le \alpha_i \le C\ \forall\ i$ and $\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4 + \alpha_5 y_5 = 0$. That is, we need $0 \le \alpha_i \le C\ \forall\ i$ and $\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 - \alpha_5 = 0$. These constraints give us a five-dimensional plane in the positive box $0 \le \alpha_i \le C\ \forall\ i$.

Let $H = \{(\alpha_1, \ldots, \alpha_4, \alpha_5) \in \mathbb{R}^5 | 0 \le \alpha_i \le C\ \forall\ i\ and\ \ \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 - \alpha_5 = 0\}$. We want to maximize $L_D(\alpha_1, \ldots, \alpha_4, \alpha_5)$ on H.

To find the maximum of $L_D(\alpha_1, \ldots, \alpha_4, \alpha_5)$ on H, we can use any computational software.

It turns out that, for $C = 2$, the maximum value of L_D on H is 6 and it occurs at $(\alpha_1, \ldots, \alpha_4, \alpha_5) = (2, 2, 2, 1, 1)$.

$$\beta = \sum_{i=1}^{5}\alpha_i y_i x_i$$

$$\Rightarrow \qquad \beta = \begin{bmatrix} -2 \\ 0 \end{bmatrix}.$$

By complementary slackness, we have $\alpha_i\big(1-\varepsilon_i - y_i(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})\big) = 0 \ \forall\ i = 1,\dots,5$ and $\mu_i\varepsilon_i = 0\ \forall\ i$.

Since $\alpha_4, \alpha_5 \neq C$ and $\alpha_i = C - \mu_i\ \forall\ i$, $\mu_4, \mu_5 \neq 0$. Thus, $\varepsilon_4 = \varepsilon_5 = 0$ since $\mu_i\varepsilon_i = 0\ \forall\ i$.

Using $\alpha_i\big(1-\varepsilon_i - y_i(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})\big) = 0\ \forall\ i$, we can solve for β_0 and the remaining ε's. We get $\beta_0 = 1, \varepsilon_1 = 0, \varepsilon_2 = 0, \varepsilon_3 = 2, \varepsilon_4 = 0, \varepsilon_5 = 0$.

The equation of our hyperplane is given by $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$.

So we get $1 - 2X_1 = 0$

$$\implies \qquad X_1 = \frac{1}{2}$$

Since $\alpha_i > 0$ for $i = 1,\dots,5$, we have that each x_i satisfies $y_i(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}) = 1 - \varepsilon_i$. Hence, x_1, x_2, x_3, x_4, x_5 are all support vectors.

b) For $C = 4$, L_D has an absolute max value of 10 and it occurs at $(\alpha_1,\dots,\alpha_5) = (4,2,4,2,0)$.

$\beta = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \beta_0 = 1, \varepsilon_1 = 0, \varepsilon_2 = 0, \varepsilon_3 = 2, \varepsilon_4 = 0, \varepsilon_5 = 0$, and the hyperplane is $X_1 = \frac{1}{2}$, the same result we got for $C = 2$.

c) For $C = 1$, L_D has an absolute max value of $\frac{7}{2}$ and it occurs at $(\alpha_1,\dots,\alpha_5) = \left(1,1,1,\frac{1}{2},\frac{1}{2}\right)$.

$\beta = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \beta_0 = 0, \varepsilon_1 = 1, \varepsilon_2 = 1, \varepsilon_3 = 1, \varepsilon_4 = 0, \varepsilon_5 = 0$, and the hyperplane is $X_1 = 0$.