Solution Set: Maximal Margin Classifier

1. Our convex optimization problem takes the form:

$\underset{(\beta_0,\beta)\in\mathbb{R}^3}{minimize} \quad f(\beta_0, \beta_1, \beta_2)$ given the constraint $g_i(\beta_0, \beta_1, \beta_2) \leq 0$ for $i = 1, 2, 3$

where $f(\beta_0, \beta_1, \beta_2) = \frac{1}{2}\|\beta\|^2$

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

So
$$g_1 = 1 - (\beta_0 + \beta_1 + \beta_2)$$
$$g_2 = 1 + (\beta_0 + 2\beta_1 + 3\beta_2)$$
$$g_3 = 1 + (\beta_0 + 3\beta_1 + \beta_2)$$

The dual Lagrangian is given by $L_D(x, \alpha) = \sum_{i=1}^{3} \alpha_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j x_i^T x_j$.

So $L_D(x, \alpha) = (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}[2\alpha_1^2 + 13\alpha_2^2 + 10\alpha_3^2 - 10\alpha_1\alpha_2 - 8\alpha_1\alpha_3 + 18\alpha_2\alpha_3]$

We want to maximize $L_D(x, \alpha)$ subject to the constraints $\alpha_i \geq 0 \; \forall \; i$ and $\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 = 0$. That is, we need $\alpha_i \geq 0 \; \forall \; i$ and $\alpha_1 - \alpha_2 - \alpha_3 = 0$. Using $\alpha_1 = \alpha_2 + \alpha_3$, rewrite L_D as follows:

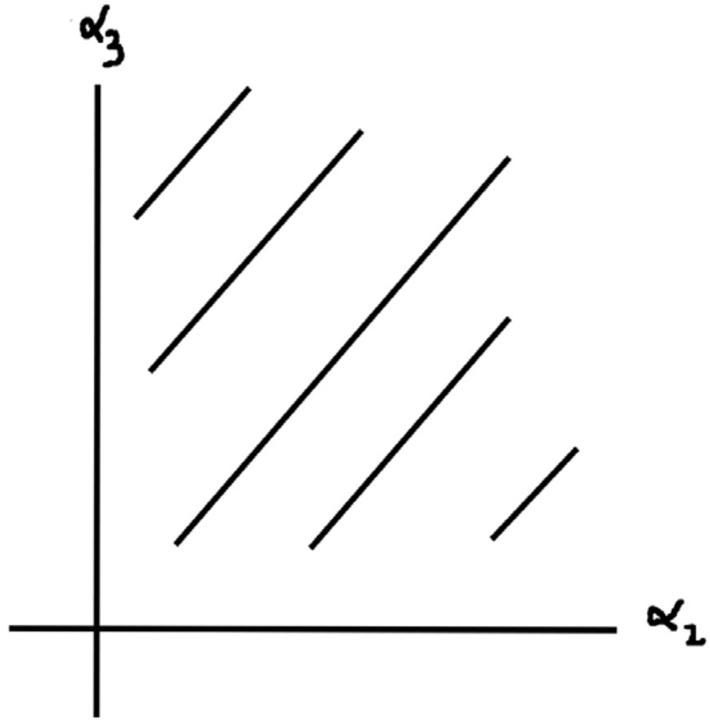
$$L_D = 2(\alpha_2 + \alpha_3) - \frac{1}{2}[2(\alpha_2 + \alpha_3)^2 + 13\alpha_2^2 + 10\alpha_3^2 - 10(\alpha_2 + \alpha_3)\alpha_2 - 8(\alpha_2 + \alpha_3)\alpha_3 + 18\alpha_2\alpha_3]$$

Simplifying, we get

$$L_D = 2(\alpha_2 + \alpha_3) - \frac{1}{2}[5\alpha_2^2 + 4\alpha_2\alpha_3 + 4\alpha_3^2]$$

So we want to maximize L_D subject to the constraints $\alpha_2 \geq 0$ and $\alpha_3 \geq 0$.

So we're maximizing L_D on the positive orthant $\alpha_2 \geq 0, \alpha_3 \geq 0$:



Let's look for any critical points in the interior of the positive orthant by setting $\nabla L_D = 0$.

$$\frac{\partial L_D}{\partial \alpha_2} = 2 - \frac{1}{2}(10\alpha_2 + 4\alpha_3) = 2 - (5\alpha_2 + 2\alpha_3)$$

$$\frac{\partial L_D}{\partial \alpha_3} = 2 - \frac{1}{2}(4\alpha_2 + 8\alpha_3) = 2 - (2\alpha_2 + 4\alpha_3)$$

Setting $\nabla L_D = 0 \quad \Longrightarrow \quad 5\alpha_2 + 2\alpha_3 = 2$

$$2\alpha_2 + 4\alpha_3 = 2$$

$$\Longrightarrow \quad \alpha_2 = \frac{1}{4} \text{ and } \alpha_3 = \frac{3}{8}$$

So $\left(\frac{1}{4}, \frac{3}{8}\right)$ is a critical point in the interior of the positive orthant.

$L_D|_{\left(\frac{1}{4}, \frac{3}{8}\right)} = \frac{5}{8}$.

Using the second derivative test, we can show that $L_D(\alpha)$ has a local max at $\left(\frac{1}{4}, \frac{3}{8}\right)$. However, a local max of a concave function on a convex set is a global max. $L_D(\alpha_2, \alpha_3)$ is a concave function and the positive orthant $E = \{(\alpha_2, \alpha_3) | \alpha_2, \alpha_3 \geq 0\}$ is convex. Hence, $L_D(\alpha_2, \alpha_3)$ has a global max on E. $L_D(\alpha_1, \alpha_2, \alpha_3)$ has a global max at $(\alpha_1, \alpha_2, \alpha_3)$ over the set

$$F = \{(\alpha_1, \alpha_2, \alpha_3) | \alpha_1 = \alpha_2 + \alpha_3, \alpha_2 \geq 0, \alpha_3 \geq 0\}$$

if and only if $L_D(\alpha_2, \alpha_3)$ has a global max at (α_2, α_3) over the set E. It follows that $L_D(\alpha_1, \alpha_2, \alpha_3)$ has a global max at $\left(\frac{5}{8}, \frac{1}{4}, \frac{3}{8}\right)$. $\left(\alpha_1 = \alpha_2 + \alpha_3 = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}\right)$

$$\beta = \sum_{i=1}^{3} \alpha_i y_i x_i = \frac{5}{8}\begin{bmatrix}1\\1\end{bmatrix} - \frac{1}{4}\begin{bmatrix}2\\3\end{bmatrix} - \frac{3}{8}\begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}-1\\ -\frac{1}{2}\end{bmatrix}$$

$\Rightarrow \quad \beta_1 = -1, \beta_2 = -\frac{1}{2}.$

By complementary slackness, $\alpha_i\left(1 - y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})\right) = 0 \ \forall\ i$

For $i = 1$, we get $\frac{5}{8}\left(1 - (\beta_0 + \beta_1 + \beta_2)\right) = 0$

$$\Rightarrow \quad 1 - \left(\beta_0 + (-1) - \frac{1}{2}\right) = 0$$

$$\Rightarrow \quad \beta_0 = \frac{5}{2}.$$

$\Rightarrow \quad \beta_0 = \frac{5}{2}$

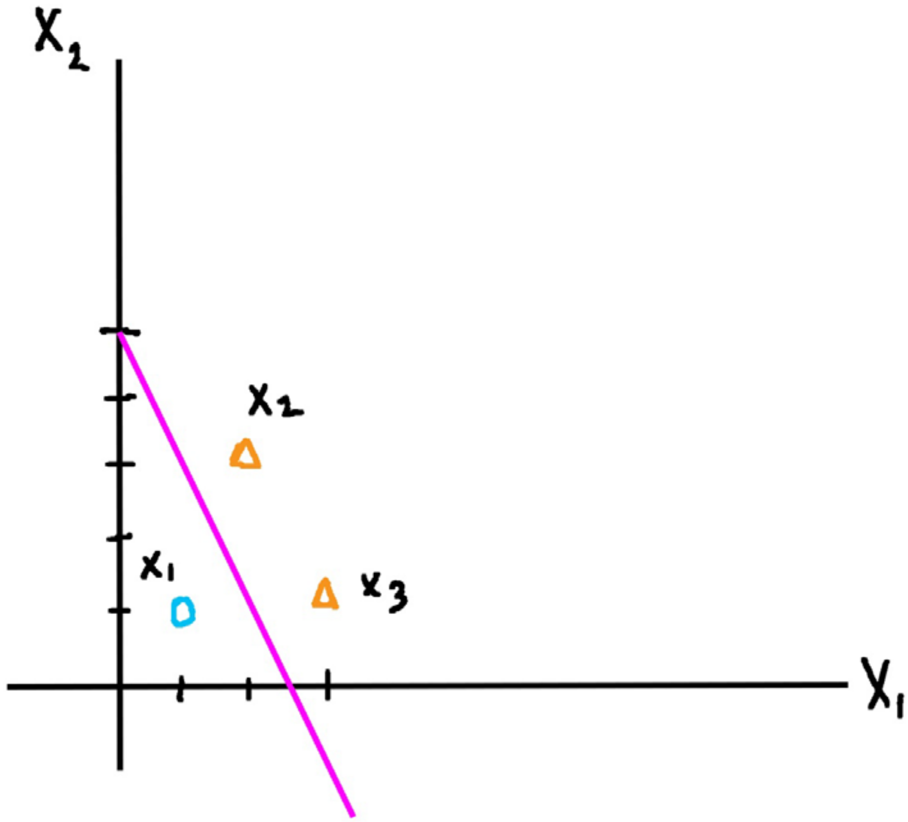
$\beta_1 = -1$

$\beta_2 = -\frac{1}{2}$

Our hyperplane is given by $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$.

So we have $\frac{5}{2} - X_1 - \frac{1}{2}X_2 = 0$

$$\Rightarrow \quad X_2 = -2X_1 + 5$$



Since $\alpha_1, \alpha_2, \alpha_3$ are all nonzero, we have that each x_i satisfies $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = 1$. Hence, x_1, x_2, x_3 all lie on the margin and are, therefore, support vectors.

2. Our convex optimization problem takes the form:

$$\underset{(\beta_0,\beta)\in\mathbb{R}^3}{minimize} \quad f(\beta_0, \beta_1, \beta_2)$$

given the constraint $g_i(\beta_0, \beta_1, \beta_2) \leq 0$ for $i = 1, 2, 3, 4$

where $f(\beta_0, \beta_1, \beta_2) = \frac{1}{2}\|\beta\|^2$ and

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

So
$$g_1 = 1 - (\beta_0 + \beta_1 + \beta_2)$$
$$g_2 = 1 + (\beta_0 + 2\beta_1 + 3\beta_2)$$
$$g_3 = 1 + (\beta_0 + 3\beta_1 + \beta_2)$$
$$g_4 = 1 - (\beta_0 + 2\beta_2)$$

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

So $L_D(x, \alpha) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) - \frac{1}{2}[2\alpha_1^2 + 13\alpha_2^2 + 10\alpha_3^2 + 4\alpha_4^2 - 10\alpha_1\alpha_2 - 8\alpha_1\alpha_3 + 4\alpha_1\alpha_4 + 18\alpha_2\alpha_3 - 12\alpha_2\alpha_4 - 4\alpha_3\alpha_4]$

We want to maximize $L_D(x,\alpha)$ subject to the constraints $\alpha_i \geq 0 \;\forall\, i$ and $\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4 = 0$. That is, we need $\alpha_i \geq 0 \;\forall\, i$ and $\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 0$. Using $\alpha_1 = \alpha_2 + \alpha_3 - \alpha_4$, rewrite L_D as follows:

$$L_D = 2(\alpha_2 + \alpha_3) - \frac{1}{2}[2(\alpha_2 + \alpha_3 - \alpha_4)^2 + 13\alpha_2^2 + 10\alpha_3^2 + 4\alpha_4^2 - 10(\alpha_2 + \alpha_3 - \alpha_4)\alpha_2 - 8(\alpha_2 + \alpha_3 - \alpha_4)\alpha_3 + 4(\alpha_2 + \alpha_3 - \alpha_4)\alpha_4 + 18\alpha_2\alpha_3 - 12\alpha_2\alpha_4 - 4\alpha_3\alpha_4]$$

Simplifying, we get

$$L_D = 2(\alpha_2 + \alpha_3) - \frac{1}{2}[5\alpha_2^2 + 4\alpha_3^2 + 2\alpha_4^2 + 4\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 4\alpha_3\alpha_4]$$

So we want to maximize L_D subject to the constraints $\alpha_2, \alpha_3, \alpha_4 \geq 0$.

So we're maximizing L_D on the positive orthant $\alpha_2, \alpha_3, \alpha_4 \geq 0$.

Let's look for any critical points in the interior of the positive orthant by setting $\nabla L_D = 0$.

$$\frac{\partial L_D}{\partial \alpha_2} = 2 - 5\alpha_2 - 2\alpha_3 - \alpha_4$$

$$\frac{\partial L_D}{\partial \alpha_3} = 2 - 4\alpha_3 - 2\alpha_2 - 2\alpha_4$$

$$\frac{\partial L_D}{\partial \alpha_4} = -2\alpha_4 + \alpha_2 - 2\alpha_3$$

Setting $\nabla L_D = 0 \quad \Longrightarrow$

$$-5\alpha_2 - 2\alpha_3 - \alpha_4 = -2$$

$$-2\alpha_2 - 4\alpha_3 - 2\alpha_4 = -2$$

$$\alpha_2 - 2\alpha_3 - 2\alpha_4 = 0$$

The solution to this system is $\alpha_2 = \frac{1}{4}, \alpha_3 = \frac{5}{8}, \alpha_4 = -\frac{1}{2}$. However, since α_4 is negative, this solution is not in the interior of the positive orthant.

We need to check the boundaries $\alpha_2 = 0, \alpha_3 = 0$, and $\alpha_4 = 0$.

On $\alpha_2 = 0$, there are no critical points in the interior of the face $\alpha_2 = 0$.

On $\alpha_3 = 0$, L_D has a local max at $(\alpha_2, \alpha_4) = \left(\frac{4}{9}, \frac{2}{9}\right)$ relative to the boundary $\alpha_3 = 0, \alpha_2 \geq 0, \alpha_4 \geq 0$. The value of L_D at $(\alpha_2, \alpha_4) = \left(\frac{4}{9}, \frac{2}{9}\right)$ is $\frac{4}{9}$.

On $\alpha_4 = 0$, L_D has a local max at $(\alpha_2, \alpha_3) = \left(\frac{1}{4}, \frac{3}{8}\right)$ relative to the boundary $\alpha_4 = 0, \alpha_2, \alpha_3 \geq 0$. The value of L_D at $(\alpha_2, \alpha_3) = \left(\frac{1}{4}, \frac{3}{8}\right)$ is $\frac{5}{8}$.

Since $\frac{5}{8} > \frac{4}{9}$, the candidate for the global max is $\left(\frac{1}{4}, \frac{3}{8}, 0\right)$. In fact, we can show that, for a fixed α_4, the local max value relative to the plane $l_{\alpha_4} = \{(\alpha_2, \alpha_3, \alpha_4) | \alpha_2, \alpha_3 \geq 0\}$ decreases as α_4 increases. (The local max occurs at $(\alpha_2, \alpha_3) = \left(\frac{1}{4}, \frac{3}{8} - \frac{\alpha_4}{2}\right)$ and $L_D = \frac{5}{8} - \frac{1}{2}(\alpha_4^2 + \alpha_4)$ there.)

Therefore, there is a global max at $\left(\frac{1}{4}, \frac{3}{8}, 0\right)$.

$L_D(\alpha_2, \alpha_3, \alpha_4)$ has a global max at $(\alpha_2, \alpha_3, \alpha_4)$ over the positive orthant $\{(\alpha_2, \alpha_3, \alpha_4) | \alpha_2, \alpha_3, \alpha_4 \geq 0\}$ if and only if $L_D(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ has a global max at $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ over the set

$$\{(\alpha_1, \alpha_2, \alpha_3, \alpha_4) | \alpha_1 = \alpha_2 + \alpha_3 - \alpha_4, \alpha_2 \geq 0, \alpha_3 \geq 0, \alpha_4 \geq 0\}.$$

It follows that $L_D(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ has a global max at $\left(\frac{5}{8}, \frac{1}{4}, \frac{3}{8}, 0\right)$.

$$\left(\alpha_1 = \alpha_2 + \alpha_3 - \alpha_4 = \frac{1}{4} + \frac{3}{8} - 0 = \frac{5}{8}\right)$$

$$\beta = \sum_{i=1}^{4} \alpha_i y_i x_i = \frac{5}{8}\begin{bmatrix}1\\1\end{bmatrix} - \frac{1}{4}\begin{bmatrix}2\\3\end{bmatrix} - \frac{3}{8}\begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}-1\\-\frac{1}{2}\end{bmatrix}$$

$\Longrightarrow \quad \beta_1 = -1, \beta_2 = -\frac{1}{2}$.

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

For $i = 1$, we get $\frac{5}{8}\big(1 - (\beta_0 + \beta_1 + \beta_2)\big) = 0$

$$\Longrightarrow \quad 1 - \left(\beta_0 + (-1) - \frac{1}{2}\right) = 0$$

$$\Longrightarrow \quad \beta_0 = \frac{5}{2}.$$

$\Longrightarrow \quad \beta_0 = \frac{5}{2}$

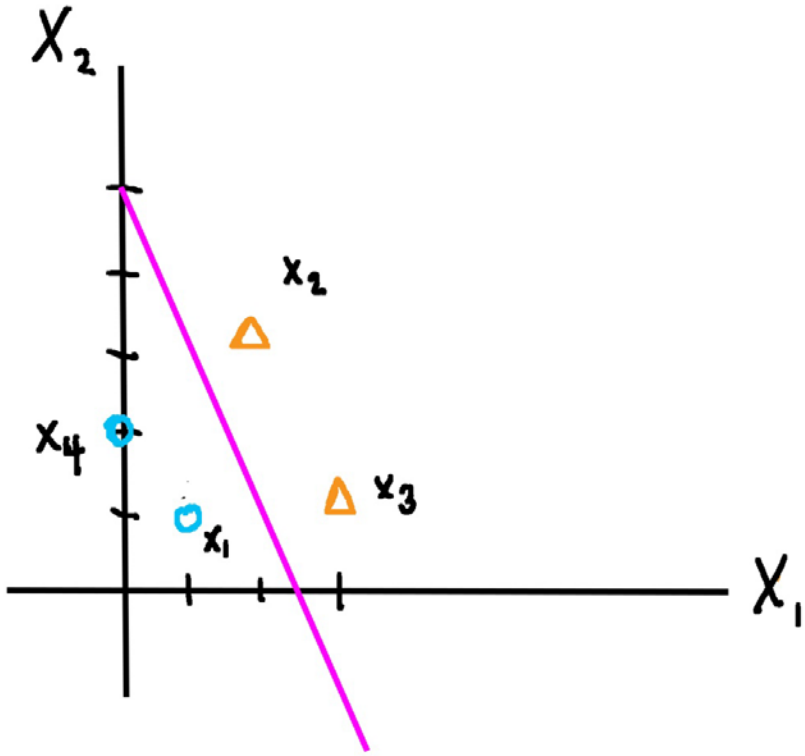
$\beta_1 = -1$

$\beta_2 = -\frac{1}{2}$

Our hyperplane is given by $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$.

So we have $\frac{5}{2} - X_1 - \frac{1}{2}X_2 = 0$

$$\Longrightarrow \quad X_2 = -2X_1 + 5$$



Since $\alpha_1, \alpha_2, \alpha_3$ are all nonzero, we have that x_1, x_2, x_3 satisfy $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = 1$. Hence, x_1, x_2, x_3 lie on the margin and are, therefore, support vectors. Notice that our hyperplane is exactly the same line we got for problem 1.