Dual Maximum Margin Optimization

Observations:

- Support vector machines can be viewed as the dual to maximum margin classifiers.
- We derive this dual representation using Lagrangian optimization theory.

Dual Maximum Margin Optimization

Assume that we are given a linearly separable training set of the following form,

$$D = \{(\overline{x}_1, y_1), (\overline{x}_2, y_2), \ldots, (\overline{x}_l, y_l)\} \subseteq \mathbb{R}^n \times \{+1, -1\},$$

then recall our primal maximum margin optimization problem,

$$\min_{\overline{w}, b} \phi(\overline{w}, b) = \min_{\overline{w}, b} \frac{1}{2} \overline{w} \bullet \overline{w},$$

subject to the constraints,

$$g_i(\overline{w}, b) = y_i(\overline{w} \bullet \overline{x}_i - b) - 1 \geq 0,$$

for $i = 1, \ldots, l$.

The constraints are rewritten in a form amenable for the Lagrangian objective function.

Dual Maximum Margin Optimization

We construct the corresponding Lagrangian as,

$$
\begin{aligned}
L(\overline{\alpha}, \overline{w}, b) &= \phi(\overline{w}, b) - \sum_{i=1}^{l} \alpha_i g_i(\overline{w}, b) \\
&= \frac{1}{2}\, \overline{w} \bullet \overline{w} - \sum_{i=1}^{l} \alpha_i (y_i(\overline{w} \bullet \overline{x}_i - b) - 1) \\
&= \frac{1}{2}\, \overline{w} \bullet \overline{w} - \sum_{i=1}^{l} \alpha_i y_i \overline{w} \bullet \overline{x}_i + b \sum_{i=1}^{l} \alpha_i y_i + \sum_{i=1}^{l} \alpha_i .
\end{aligned}
$$

This gives us the Lagrangian optimization problem for maximum margin classifiers,

$$
\max_{\overline{\alpha}} \min_{\overline{w}, b} L(\overline{\alpha}, \overline{w}, b),
$$

subject to,

$$
\alpha_i \geq 0,
$$

for $i = 1, \ldots, l$.

Dual Maximum Margin Optimization

Now, let $\overline{\alpha}^*$, $\overline{w}^*$, and b^* be a solution to the Lagrangian optimization problem such that,

$$\max_{\overline{\alpha}} \min_{\overline{w},b} L(\overline{\alpha}, \overline{w}, b) = L(\overline{\alpha}^*, \overline{w}^*, b^*).$$

But,

- ϕ is convex
- the constraints g_i are linear

this implies that the solution $\overline{\alpha}^*$, $\overline{w}^*$ and b^* will satisfy the following KKT conditions,

$$\begin{aligned}
\frac{\partial L}{\partial \overline{w}}(\overline{\alpha}^*, \overline{w}^*, b^*) &= \overline{0}, \\
\frac{\partial L}{\partial b}(\overline{\alpha}^*, \overline{w}^*, b^*) &= 0, \\
\alpha_i^*(y_i(\overline{w}^* \bullet \overline{x}_i - b^*) - 1) &= 0, \\
y_i(\overline{w}^* \bullet \overline{x}_i - b^*) - 1 &\geq 0, \\
\alpha_i^* &\geq 0,
\end{aligned}$$

for $i = 1, \ldots, l$.

Dual Maximum Margin Optimization

Of particular interest is of course the third condition,

$$\alpha_i^*(y_i(\overline{w}^* \bullet \overline{x}_i - b^*) - 1) = 0,$$

the complimentarity condition, because it assures the existence of a solution to our primal maximum margin optimization problem,

$$\begin{aligned}
\max_{\overline{\alpha}} \min_{\overline{w},b} L(\overline{\alpha}, \overline{w}, b) = L(\overline{\alpha}^*, \overline{w}^*, b^*) \\
&= \frac{1}{2}\overline{w}^* \bullet \overline{w}^* - \sum_{i=1}^{l} \alpha_i^*(y_i(\overline{w}^* \bullet \overline{x}_i - b^*) - 1) \\
&= \frac{1}{2}\overline{w}^* \bullet \overline{w}^* \\
&= \phi(\overline{w}^*, b^*).
\end{aligned}$$

Dual Maximum Margin Optimization

As in our simple, one dimensional example we want to solve the Lagrangian optimization problem by constructing and solving the Lagrangian dual.

In order to accomplish this we first construct expressions for the optima of the primal variables at the saddle point. We know that they exist because of the KKT conditions. Therefore,

$$\frac{\partial L}{\partial \overline{w}}(\overline{\alpha}, \overline{w}^*, b) = \overline{w}^* - \sum_{i=1}^{l} \alpha_i y_i \overline{x}_i = \overline{0}.$$

It follows that,

$$\overline{w}^* = \sum_{i=1}^{l} \alpha_i y_i \overline{x}_i.$$

And also,

$$\frac{\partial L}{\partial b}(\overline{\alpha}, \overline{w}, b^*) = \sum_{i=1}^{l} \alpha_i y_i = 0,$$

giving rise to the constraint

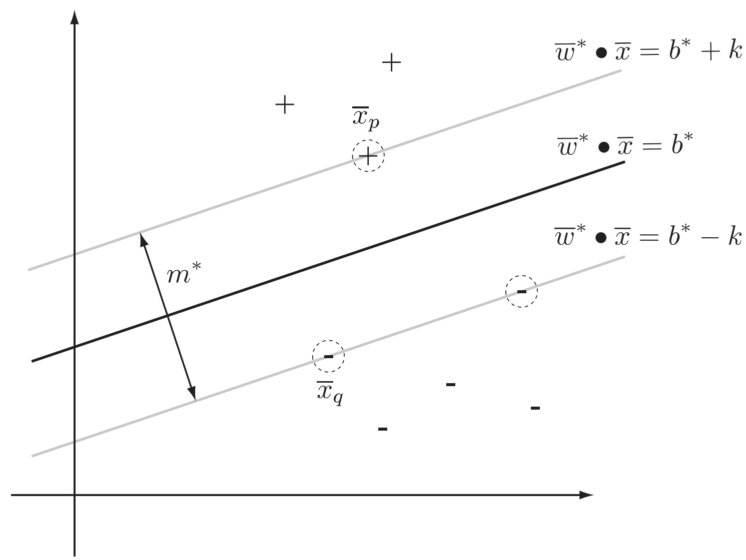
$$\sum_{i=1}^{l} \alpha_i y_i = 0,$$

that needs to hold at the saddle point.

Dual Maximum Margin Optimization

Observation: No expression for b^*!

However, we can recover an expression for b^* from the structure of the data set.



$$b^+ = \overline{w}^* \bullet \overline{x}_p$$

$$b^- = \overline{w}^* \bullet \overline{x}_q$$

or computationally,

$$b^+ = \min\{\overline{w}^* \bullet \overline{x} \mid (\overline{x}, y) \in D \text{ with } y = +1\}$$

again computationally,

$$b^- = \max\{\overline{w}^* \bullet \overline{x} \mid (\overline{x}, y) \in D \text{ with } y = -1\}$$

Dual Maximum Margin Optimization

Now, the decision surface with b^* as an offset sits right in the middle of the margin between the two supporting hyperplanes, therefore

$$b^* = \frac{b^+ + b^-}{2}.$$

Dual Maximum Margin Optimization

We are now ready to construct our Lagrangian dual,

$$\phi'(\overline{\alpha}) = L(\overline{\alpha}, \overline{w}^*, b^*) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \overline{x}_i \bullet \overline{x}_j,$$

by applying the identity for $\overline{w}^*$ and the newly found constraint to the Lagrangian.

Dual Maximum Margin Optimization

Proposition: (The Maximum Margin Lagrangian Dual) Given the primal maximum margin optimization, [a] then the Lagrangian dual optimization for maximum margin classifiers is

$$\max_{\overline{\alpha}} \phi'(\overline{\alpha}) = \max_{\overline{\alpha}} \left(\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \overline{x}_i \bullet \overline{x}_j \right),$$

subject to the constraints

$$\sum_{i=1}^{l} \alpha_i y_i = 0,$$
$$\alpha_i \geq 0,$$

with $i = 1, \ldots, l$.

[a] See lecture notes on maximum margin classifiers.

Dual Maximum Margin Optimization

Given a solution $\overline{\alpha}^*$ to the Lagrangian dual optimization, then the KKT complementarity condition can only be satisfied for each $i = 1, \ldots, l$ if either $\alpha_i^* = 0$ or $y_i(\overline{w}^* \bullet \overline{x}_i - b^*) - 1 = 0$.

If we consider $\alpha_j^* > 0$ for some point $(\overline{x}_j, y_j) \in D$, then in order to satisfy the complementarity condition we have $y_j(\overline{w}^* \bullet \overline{x}_j - b^*) - 1 = 0$ or,

$$
\begin{aligned}
\overline{w}^* \bullet \overline{x}_j &= b^* + 1 \quad \text{if } y_j = +1, \\
\overline{w}^* \bullet \overline{x}_j &= b^* - 1 \quad \text{if } y_j = -1.
\end{aligned}
$$

That is, the point $(\overline{x}_j, y_j)$ lies on one of the supporting hyperplanes! It is a constraint!

Now consider $\alpha_j^* = 0$ for some point $(\overline{x}_j, y_j) \in D$. That is, the point $\overline{x}_j$ is a point that does not lie in the vicinity of the class boundary because we have $y_j(\overline{w}^* \bullet \overline{x}_j - b^*) - 1 > 0$ or,

$$
\begin{aligned}
\overline{w}^* \bullet \overline{x}_j &> b^* + 1 \quad \text{if } y_j = +1, \\
\overline{w}^* \bullet \overline{x}_j &< b^* - 1 \quad \text{if } y_j = -1.
\end{aligned}
$$

This implies that points with zero-valued Lagrangian multipliers do not constrain the size of the margin.

Points with non-zero Lagrangian multipliers are support vectors!

Dual Maximum Margin Optimization

This gives us the following insight,

The primal maximum margin optimization computes the supporting hyperplanes whose margin is limited by support vectors. The dual maximum margin optimization computes the support vectors that limit the size of the margin of the supporting hyperplanes.