CS281B/Stat241B: Advanced Topics in Learning & Decision Making

Lagrangian Duality

Lecturer: Michael I. Jordan *Scribes: Peter Kassakian*

Recall that we wish to build an optimal margin classifier. This problem is known to be a quadratic programming problem (QP), under convex constraints. The problem is to find the hyperplane that separates the labelled data that has the maximal margin (distance from the separating hyperplane to the nearest data points). This is formulated as follows:

$$\begin{aligned} \min \quad w^T w \qquad \text{such that} \quad & x_i^T w + b \geq 1, \quad x_i \quad \text{in class} \quad 1 \\ & x_i^T w + b \leq -1, \quad x_i \quad \text{in class} \quad -1 \end{aligned} \tag{1}$$

Letting $y_i \in -1, 1$ be the label of x_i, we can rewrite Equation 1 compactly as:

$$\min \quad w^T w \quad \text{such that} \quad y_i(x_i^T w + b) \geq 1 \tag{2}$$

This problem could be handed off to an optimization routine at this point, but we would lose valuable insight if we took that route.

1 Lagrangian Duality

We will attack this problem by looking at the *primal* and *dual* problems. Assume we have a problem of the form:

$$\begin{aligned} \min \quad f(x) \qquad \text{such that} \qquad & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in X \text{convex} \end{aligned} \tag{3}$$

Of course the problem above (Equation 2) is of this form. We can solve this using Lagrange multipliers. The Lagrangian is found by multiplying the constraints by Lagrange multipliers (scalars) and adding them to the expression to be minimized:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda g(x) + \mu h(x) \tag{4}$$

Taking derivatives of the Lagrangian with respect to λ and μ, and setting to zero yields the constraint equations. Setting all the derivatives of $\mathcal{L}(x, \lambda, \mu)$ to zero, and solving will yield an interesting point. What type of point is it that we're interested in? Is it a saddle, minimum, maximum, or what? It turns out that we're interested in *minimizing* with respect to x, and *maximizing* with respect to λ and μ. Why?

Equation 3 is a formulation of the primal problem. For every primal problem, there is a dual problem that might be easier to solve. A major advantage to solving the dual problem in lieu of the primal problem is that the dual problem is always concave. This advantage comes at the expense that solving the dual problem might yield a lower bound instead of an optimum. Under certain conditions this lower bound is indeed the optimum, however. The dual problem is formulated as follows:

$$\max_{u,v} \theta(u,v), \quad \text{where} \quad \theta(u,v) = \min_x \{f(x) + u^T g(x) + v^T h(x)\}, \quad \text{such that} \quad u \geq 0 \tag{5}$$

Here $f(x)+u^Tg(x)+v^Th(x)$ is the Lagrangian $\mathcal{L}(x,u,v)$. So no matter what $f()$ is, solving the dual problem is always concave. Note that it is not always easy to find a closed form expression for the dual problem, however.

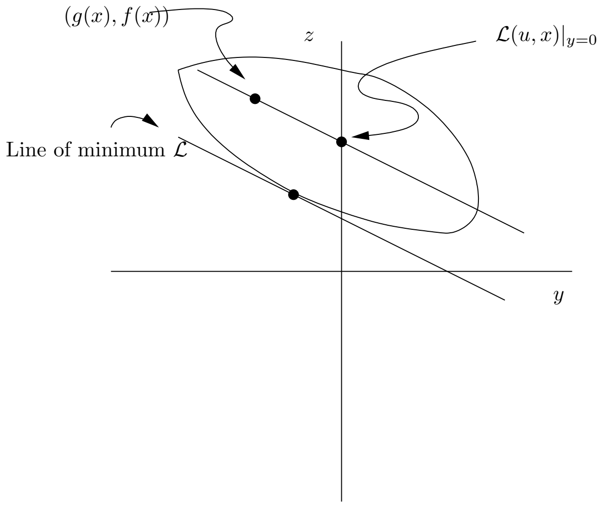
Ignoring $h(x)$ for simplicity, Let's look at what's happening graphically.

$$\text{Let} \quad G = \{(y,z) : y = g(x), z = f(x) \quad \text{for some} \quad x \in X\} \tag{6}$$

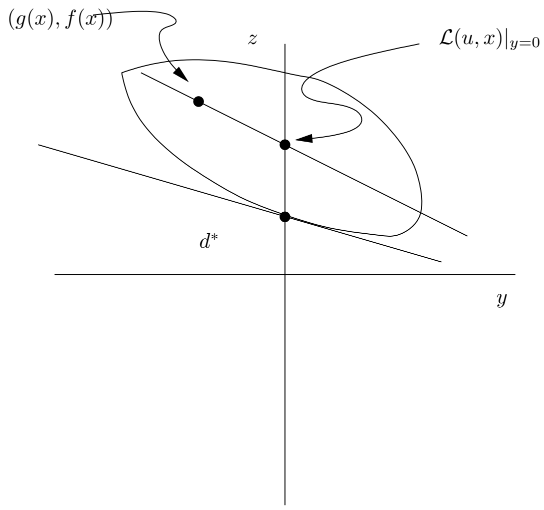
In Figure 1, the enclosed region is the set of all points $(g(x), f(x))$ for $x \in X$. Our formulation of the dual problem can be written:

$$\max_{u,v} \min_x \mathcal{L}(x,u) \quad \text{such that} \quad u \geq 0 \tag{7}$$

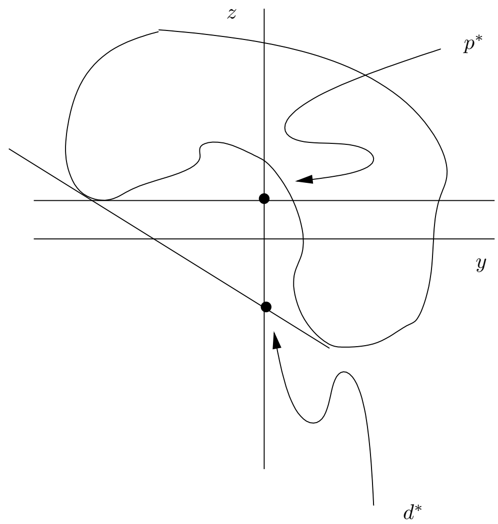
To perform this max min operation, we can first keep u fixed, and minimize $\mathcal{L}$ with respect to x. Note that in Figure 1, points of fixed $\mathcal{L}$ for fixed u are lines with slope $-u$. The z-intercept, corresponding to evaluating $f(x)$ when $g(x) = 0$, is the value of $\mathcal{L}(x,u)$ for any $(g(x), f(x))$ on lines of a given slope $-u$. This is all a way of saying that you minimize $\mathcal{L}(x,u)$ for fixed u, by shifting the lines of constant $\mathcal{L}(x,u)$ *down*. Minimizing with respect to x will find lines that are tangent to the *bottom* of the set of all possible points $(g(x), f(x))$. See Figure 1.



The second part of the optimization of the dual problem, is to maximize $\mathcal{L}$ with respect to u. In Figure 1, since the value of $\mathcal{L}$ can be seen as the z-intercept, we notice that of all the lines tangent to the bottom of the region of interest, the one that has the greatest z-intercept, and therefore the largest value of $\mathcal{L}$, is the line that is tangent right at its z-intercept. See Figure 1.



In Figure 1, d^* is the optimal point for the dual problem. Thinking again about the primal problem, we realize that minimizing $f(x)$ such that $g(x) \leq 0$ is equivalent to finding the point in G and also in the left half plane that has the smallest value of z. So, we can see that in the case that the region of interest is convex in the left half plane, the optimal point found in the dual problem is the same as that of the primal problem. Figure 1 shows an example of a case where the solving the dual problem only yields a lower bound. There is what's known as a *duality gap*.



There will be no duality gap if $f(x)$ and $g(x)$ are convex and a contraint qualification is satisfied (e.g., the Slater condition).

Looking at this situation from another point of view, can see that the maximization of $\mathcal{L}(x,u)$ with respect to u when $u \geq 0$ is ∞ if $g(x) > 0$ since u can be any positive number. If $g(x) \leq 0$, then the value of u which makes $\mathcal{L}(x,u)$ a maximum is 0, since the $ug(x)$ term would only contribute a nonpositive number. This implies:

$$\max_{u, u \geq 0} \mathcal{L}(x,u) = f(x) \quad \text{if} \quad g(x) \geq 0 \tag{8}$$

So by substitution of this representation of $f(x)$ into the original primal problem, we see that $\min_x \max_u \mathcal{L}(x,u)$ solves the primal problem.

This looks similar to the dual problem. What if $\min_x \max_u \mathcal{L}(x,u) == \max_u \min_x \mathcal{L}(x,u)$? If this were the case, then the right hand side would solve the primal problem. We take note that the right hand side is exactly our dual problem formulation.

The following is always true:

$$\min_x \max_u \mathcal{L}(x,u) \geq \max_x \min_u \mathcal{L}(x,u) \tag{9}$$

for any function $\mathcal{L}$. Equation 9 is a statement of *weak duality.* Equality will hold if and only if (x^*, u^*) is a saddle point of $\mathcal{L}(x,u)$.

Theorem: There is no duality gap iff $\mathcal{L}$ has a saddle point.

Theorem: $\mathcal{L}$ having a saddle point implies the KKT conditions.