Machine Learning (CS 5350/CS 6350) *29 Jan 2008*

Optimization theory

This is a brief, simplified "review" of optimization theory. This gives roughly the right ideas, but in order to actually prove some of these results, you need to make stronger assumptions (sketched at the end).

Suppose we want to solve the following optimization problem:

$$\begin{array}{rl} \text{minimize}_{\boldsymbol{w}} & f(\boldsymbol{w}) \\ \text{subject to} & g_n(\boldsymbol{w}) \leq 0 \qquad , \quad n = 1, \dots, N \end{array}$$

In order to solve this, we form the *Lagrangian dual*:

$$L(\boldsymbol{w}, \boldsymbol{\alpha}) = f(\boldsymbol{w}) + \sum_n \alpha_n g_n(\boldsymbol{w})$$

It is useful to also define the inf-dual (non-standard terminology) as:

$$\hat{L}(\boldsymbol{\alpha}) = \inf_{\boldsymbol{w}} f(\boldsymbol{w}) + \sum_n \alpha_n g_n(\boldsymbol{w})$$

You should think of inf as roughly the same as min. Given the dual functions, we define the *dual problem* (the original is called the "primal problem") as:

$$\begin{array}{rl} \text{maximize}_{\boldsymbol{\alpha}} & \hat{L}(\boldsymbol{\alpha}) \\ \text{subject to} & \alpha_n \geq 0 \qquad , \quad n = 1, \dots, N \end{array}$$

The hope is that this is easier to solve than the primal (essentially because the constraints are "simple"). Note also that the minimization has turned into a maximization.

Under appropriate conditions, the Strong Duality Theorem states that solving the primal and dual problems are equivalent. In other words, if it's *easier* for us to solve the dual, then we might as well solve the dual because the answer will be the same. (Note that all that's guaranteed is that the *value* is the same; since the primal is an optimization over $\boldsymbol{w}$ and the dual is an optimization over $\boldsymbol{\alpha}$, solving the dual will only give us an answer if we can express $\boldsymbol{w}$ in terms of $\boldsymbol{\alpha}$. Fortunately, this is always possible for our problems.)

However, a stronger result, known as the Kuhn-Tucker theorem, gives us more to work with. Kuhn-Tucker states that there are the following conditions that are both necessary and sufficient for $\boldsymbol{w}^*$ to be an optimal solution to the primal. Namely, that there exists a vector $\boldsymbol{\alpha}^*$ such that:

1. $\frac{\partial L(\boldsymbol{w}^*, \boldsymbol{\alpha}^*)}{\partial \boldsymbol{w}} = 0$
2. $\alpha_n^* g_n(\boldsymbol{w}^*) = 0, n = 1, \dots, N$
3. $g_n(\boldsymbol{w}^*) \leq 0, n = 1, \dots, N$
4. $\alpha_n^* \geq 0, n = 1, \dots, N$

This conditions are sometimes referred to as the Karush-Kuhn-Tucker (or KKT) conditions. One important aspect of the KKT conditions is that they say that *if* a constraint g_n is "inactive" (i.e., it is not affecting the solution) then $\alpha_n^* = 0$ (we get this by seeing that each g_n has to be non-positive; if it is strictly negative then in order to satisfy (2), α_n^* must be zero).

The primal-dual scheme is very powerful. We will primarily use it by following the following recipe:

1. Compute $\hat{L}(\boldsymbol{\alpha})$ by differentiating $\boldsymbol{L}$ with respect to $\boldsymbol{w}$ to solve the inf problem.
2. Re-express $\boldsymbol{w}$ in terms of $\boldsymbol{\alpha}$
3. Solve the corresponding $\hat{L}$ as a function of $\boldsymbol{\alpha}$