This document gives an example of Lagrange multipliers and the Lagrangian.

1 Example problem from class

Consider trying to fence in as large an area as possible using 12 units of fencing and with the help of a cliff on one side that doesn’t need to be fenced. What width \( w \) and height \( h \) maximizes the enclosed area (see Figure 1).

![Figure 1: Enclosing 3 sides with a fence. Enclosed area is \( hw \) and fence length is \( 2h + w \).](image)

We can write this problem as:

\[
\max_{h,w} \ hw \quad \text{subject to} \quad 2h + w = 12. \tag{1}
\]

One way to solve it is to see that the constraint implies \( w = 12 - 2h \) and remove the \( w \). This turns the problem into \( \max h(12 - 2h) \) which can be solved using calculus. Setting the derivative to 0, we get \( 4h = 12 \), or \( h = 3 \) (and thus \( w = 6 \) and the enclosed area is 18).

If you can’t get such nice implications from the constraint, you can use the method of Lagrange multipliers. First notice that problem (1) is the same as:

\[
\min_{h,w} \ -hw \quad \text{subject to} \quad 2h + w - 12 = 0. \tag{2}
\]

and this is the same as:

\[
\min_{h,w} \max_{\alpha} \ -hw + \alpha(2h + w - 12). \tag{3}
\]

The equivalence of problem (3) can be seen by considering feasible and infeasible pairs \( h, w \). When an \( h, w \) pair is feasible (satisfies the constraints, \( 2h + w - 12 = 0 \) in this case) we are minimizing the objective function \( -hw \) in this case). When an \( h, w \) pair is not feasible, then \( 2h + w - 12 \) is not zero, and maximizing over \( \alpha \) gives infinity (which can’t be the minimum).

Note that if we had more constraints, then we would have additional Lagrange multipliers (one per constraint). Also, if some of the constraints were inequalities (like \( 2h + w - 12 \leq 0 \)) then the Langrange multiplier for that constraint would be constrained to be non-negative.
The inside of problem (3) is called the Lagrangian, \( L(h, w, \alpha) = -hw + \alpha(2h + w - 12) \), and \( \alpha \) is the Lagrange multiplier for the constraint.

A necessary condition for the the solution of problem (3) is that the partial derivatives of the Lagrangian (with respect to \( h \), \( w \), and \( \alpha \)) are all zero. This gives us three equations in three unknowns. In this example we get the partials:

\[
\frac{\partial L(h, w, \alpha)}{\partial h} = -w + 2\alpha \\
\frac{\partial L(h, w, \alpha)}{\partial w} = -h + \alpha \\
\frac{\partial L(h, w, \alpha)}{\partial \alpha} = 2h + w - 12.
\]

And setting them equal to zero gives:

\[
w = 2\alpha \\
h = \alpha \\
2h + w = 12.
\]

Substituting the first two into the third equation gives \( 4\alpha = 12 \), or \( \alpha = 3 \). Therefore the optimizing \( h \) and \( w \) are 3 and 6 respectively.

Where I went wrong in lecture, was re-writing \( w \) and \( h \) in terms of \( \alpha \) before taking the derivative with respect to \( \alpha \).

Now consider the dual problem:

\[
\max_{\alpha} \min_{h, w} -hw + \alpha(2h + w - 12).
\]

Moving the max outside can only reduce the value of an optimal solution. Thus the value of the dual problem is a lower bound on the value of the original primal problem (either problem (1) or (3) ). In this context the Lagrange multipliers (just \( \alpha \) in this case) are called the dual variables and the original variables (here \( h \) and \( w \)) are called the primal variables. When Strong Duality holds, the optimum values (e.g. area fenced in for this example) of the primal and dual problems are the same, and solutions for one can be used to find solutions to the other.

However, for this example, strong duality does not hold. Consider the inside of the dual:

\[
\min_{h, w} -hw + \alpha(2h + w - 12).
\]

No matter what \( \alpha \) is chosen, we can make \( h \) and \( w \) both super big so that \( hw \gg \alpha(2h + w - 12) \). Therefore the value of \( -hw + \alpha(2h + w - 12) \) can be made as small as desired, and the minimum value over all \( h, w \) pairs is conventionally taken as \( -\infty \). Therefore for this problem, strong duality does not hold.

Fortunately, strong duality does hold for the kinds of min / max problems arising with support vector machines. Furthermore, the inside of the dual (analogous to (11)) can be optimized by setting the derivatives with respect to the primal variables to zero. (Note: this is exactly the mistake I made in class – it wouldn’t have been a mistake if strong duality held for the sample problem.) This often leads to a simpler dual problem with just the dual variables remaining.
2 A second example

Consider three vegetable farmers. One farmer has an amount of fencing that he needs to sell to the other two, but he wants to minimize the area the others can plant (and thus his competition). If he has to sell 40 total units of fencing to pay his debts, how much should he sell to each other farmer? Assume that length \( \ell \) of fencing sold to another farmer will be used to enclose a square plot of size \( (\ell/4)^2 \). (You might immediately see that splitting the fence evenly minimizes the area the other two farmers can enclose at 25 square units each, for a total of 50 square units).

We can set this up as an optimization problem, with \( \ell_1 \) and \( \ell_2 \) being the primal variables for the lengths sold to the two other farmers.

\[
\min_{\ell_1, \ell_2} \frac{\ell_1^2 + \ell_2^2}{4} \quad \text{subject to} \quad \ell_1 + \ell_2 = 40. \tag{12}
\]

Using a Lagrange multiplier for the constraint, we can re-write the problem as:

\[
\min_{\ell_1, \ell_2} \max_\alpha \frac{\ell_1^2 + \ell_2^2}{4} + \alpha(\ell_1 + \ell_2 - 40). \tag{13}
\]

The dual problem is:

\[
\max_\alpha \min_{\ell_1, \ell_2} \frac{\ell_1^2 + \ell_2^2}{4} + \alpha(\ell_1 + \ell_2 - 40). \tag{14}
\]

This Lagrangian \( L(\ell_1, \ell_2, \alpha) = \frac{\ell_1^2 + \ell_2^2}{4} + \alpha(\ell_1 + \ell_2 - 40) \) has the right convexity properties for strong duality to hold (whereas the three-sided fence example did not). Taking partial derivatives with respect to the primal variables, we get:

\[
\frac{\partial L(\ell_1, \ell_2, \alpha)}{\partial \ell_1} = \frac{\ell_1}{2} + \alpha \quad \text{so} \quad \ell_1 = -2\alpha, \tag{15}
\]

\[
\frac{\partial L(\ell_1, \ell_2, \alpha)}{\partial \ell_2} = \frac{\ell_2}{2} + \alpha \quad \text{so} \quad \ell_2 = -2\alpha. \tag{16}
\]

Since \( \ell_1 = \ell_2 \), we could use the constraint (which is \( \partial L(\ell_1, \ell_2, \alpha)/\partial \alpha \)) to see that \( \ell_1 = \ell_2 = 20 \) at this point. However, in this case we can also substitute into the dual to get the problem:

\[
\max_\alpha \frac{(-2\alpha)^2 + (-2\alpha)^2}{4} + \alpha(-2\alpha - 2\alpha - 40), \tag{17}
\]

or after simplifying

\[
\max_\alpha -2\alpha^2 - 40\alpha \tag{18}
\]

This last can be solved with calculus (note that the second derivative is negative). The derivative is \(-4\alpha - 40\) which when set to zero implies the maximum is at \( \alpha = -10 \). Thus using the solutions for \( \ell_1 \) and \( \ell_2 \) in terms of \( \alpha \) on lines (15) and (16) we get \( \ell_1 = \ell_2 = 20 \) as expected.