Objectives

- The Objective Function (Primal problem)

- The Dual Problem (with Lagrange multipliers)

- Computing the gradient

- Practice Problems

1 The Objective Function

This continues a lecture on the support vector machines. Our goal is to optimize the objective function. This is also known as the primal problem.

Our objective: How do we minimize over w, b

, subject to the below constraint equation?

$$y_i(w^T x_i + b)) \geqslant 1$$

There are n constraints associated with n separate constraint equations.

Where n is the size of the feature space D.

It is possible to optimize this using Newtonian math, but this is difficult and requires the Hessian matrix:

$$\begin{bmatrix} \frac{\partial L^2}{\partial w_1 \partial w_1} & \frac{\partial L^2}{\partial w_1 \partial w_2} & \cdots & \frac{\partial L^2}{\partial w_1 \partial w_m} \\ \frac{\partial L^2}{\partial w_1 \partial w_1} & \frac{\partial L^2}{\partial w_2 \partial w_2} & \cdots & \frac{\partial L^2}{\partial w_2 \partial w_m} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial L^2}{\partial w_n \partial w_1} & \frac{\partial L^2}{\partial w_n \partial w_2} & \cdots & \frac{\partial L^2}{\partial w_n \partial w_m} \end{bmatrix}$$

2 The Dual Problem

This is tough to compute, so instead we use **Lagrange Multipliers**.

In other words, instead of solving the primal problem, we solve the]**dual problem**.

This is the same as saying that we minimnize the loss function using partial derivatives. Let's recall the loss function:

$$L = \frac{1}{2}||w||^2 - \sum \alpha_i(y_i(w^T x_i + b) - 1)$$

3 Computing the Gradient

Let's compute the derivative with respect to w, b.

$L = \frac{1}{2}||w||^2 - \sum \alpha_i(y_i(w^T x_i + b) - 1)$

3.1 The left-hand derivative

The derivative of $\frac{1}{2}||w||^2$ is relatively easy to compute. It ends up being w. Why?

$LHS = \frac{1}{2}||w||^2 = \sqrt{w_1^2 + w_2^2 + ... + w_n^2}^2 = w_1^2 + w_2^2 + ... + w_n^2$

$= \frac{\partial \frac{1}{2}||w||^2}{\partial w} = (\frac{1}{2})(2w_1 + 2w_2 + ... + 2w_n) = w$

3.2 The right-hand derivative

Let's take the derivative of the right-hand side of the equation. We will subtract it later, but for now we will omit the negative sign following the left-hand-side.

$$RHS = \sum_{i=1}^{n} \alpha_i(y_i(w^T x_i + b) - 1)$$

$$= \sum_{i=1}^{n} [\alpha_i y_i(w^T x_i + b) - \alpha_i]$$

$$= \sum_{i=1}^{n} (\alpha_i y_i w^T x_i + \alpha_i y_i b - \alpha_i)$$

$$= \sum_{i=1}^{n} \alpha_i y_i w^T x_i + \sum_{i=1}^{n} \alpha_i y_i b - \sum_{i=1}^{n} \alpha_i$$

We can now take the derivative with respect to w and b.

The derivative with respect to w

We can remove the second and third components of the equation because they do not contain w. It reduces to the following:

$$\frac{\partial \sum_{i=1}^{n} \alpha_i y_i w^T x_i}{\partial w} = \sum_{i=1}^{n} \alpha_i y_i x_i$$

Why is this true?

$$w^T x_i = w_1 x_{i1} + w_2 x_{i2} + ... + w_m x_{im}$$

$$\frac{\partial w^T x_i}{\partial w} = x_{i1} + x_{i2} + ... + x_{im} = x_i$$

The derivative with respect to b

$$\frac{\partial \sum_{i=1}^{n} \alpha_i y_i b}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i$$

because

$$\alpha_i y_i b = \alpha_i y_{i1} b + \alpha_i y_{i2} b + ... + \alpha_i y_{im} b$$

and the derivative is

$$\alpha_i y_{i1} + \alpha_i y_{i2} + ... + \alpha_i y_{im}$$

So in summary, let's recap the loss equation and the partial derivatives.

$$L = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T x^i + b) - 1)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i$$

We can then set this equal to zero and solve.

$$w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

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

Let's now recall our original loss equation:

$$L = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i(y_i(w^T x^i + b) - 1)$$

Recall that $||w||^2 = w^T w$

$$= \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i y_i w^T x^i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i)$$

$$= \frac{1}{2} w^T w - w^T \sum_{i=1}^{n} \alpha_i y_i x^i - b \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i$$

Using what we found in our derivative, we can set part of this equation equal to zero.

$$= \frac{1}{2} w^T w - w^T \sum_{i=1}^{n} \alpha_i y_i x^i + \sum_{i=1}^{n} \alpha_i$$

Recall that $w = \sum_{i=1}^{n} \alpha_i y_i x^i$

$$= \frac{1}{2} w^T w - w^T w + \sum_{i=1}^{n} \alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} w^T w$$

$$= \sum_{i=1}^{n} \alpha_i (\sum_{i=1}^{n} \alpha_i y_i x_i)^T (\sum_{i=j}^{n} \alpha_j y_j x_j)$$

XPL1:
Demonstrate that
$$\sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \beta_j = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j$$

This process leaves us with a new problem: a maximization problem, for the following equation:

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

4   Summary

In this section, we developed the minimization of the loss function for our support vector machine. In machine learning, much of what we do will involve gradient descent, where the gradient is another way of talking about the derivative of our w and b terms. Solving the loss function using the gradient is a common theme in all machine learning applications. In developing the loss function, we turned our minimization problem into a maximization problem, which will provide us more traction in getting to the solution.