Notes - Octoboer 25

Philip Warton

October 25, 2021

1 Hard-Margin Support Vector Machine

Linear decision rule:

$$\begin{aligned} \boldsymbol{w}^T\boldsymbol{x} + b > 0 &\qquad \Rightarrow 1 \\ \boldsymbol{w}^T\boldsymbol{x} + b < 0 &\qquad \Rightarrow 0 \end{aligned}$$

We solve for an optimal $\boldsymbol{w}$ and b by doing

$$\min_{\boldsymbol{w},b} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \quad : \quad y_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) \geq 1 \forall i$$

QP-solvers (Quadratic Problem Solvers) solve equations of the form

$$\min_{z} \frac{1}{2}\boldsymbol{z}^T P\boldsymbol{z} + \boldsymbol{q}^T\boldsymbol{z} \quad : \quad Gz \leq h, Az = b, etc..$$

Obviously using a QP-solver will help us find an optimal $\boldsymbol{w}, b$, so we need not implement this ourselves.

Notice that

$$\begin{bmatrix} w_0 & w_1 & b \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ b \end{bmatrix} = \begin{bmatrix} w_0 \\ w_1 \\ 0 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ b \end{bmatrix} = w^T w$$

So we let $P = \begin{bmatrix} I_d & 0 \\ 0 & 0 \end{bmatrix}_{d+1 \times d+1}$ in order to satisfy our needs.

$$y_i w^T x_i + y_i b \geq 1$$

$$\implies \begin{bmatrix} y_1 x_1^T y_1 \\ y_2 x_2^T y_2 \\ \vdots \\ y_n x_n^T y_n \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

So what we just looked at was the "Hard-Margin SVM Primal", but there is a more complicated version called the "Hard-Margin SVM Dual". However, they both end up being quadratic programs with linear constraints.

Notice the following:

- The optimal weight vector is a linear combination of only a few of our training examples

$$w^* = \sum_i \alpha_i^* y_i x_i$$

- Both the objective and classifying new examples are just based on dot-products between input vectors.

Two opposing goals

- Very large margin, many errors.
- Very small margin, few errors.

This yields a hyperparameter that we may find ourselves tweaking in homework at some point. Let's allow each example to violate our requirement by some value. That is,

$$y_i w^T x_i + y_i b \geq 1 + \eta$$

Where η is our error or "slack variable".