05 - Max Margin Classifiers

Ethan Graham

20th June 2023

Reformulating Signed Distance again

We now formulate the signed distance as being $\tilde{\mathbf{w}}' \cdot \tilde{\mathbf{x}} = \frac{\tilde{\mathbf{w}}' \cdot \tilde{\mathbf{x}}}{||\mathbf{w}||}$

The max margin classifier tries to maximize the **unsigned** distances of the closest point to the boundary.

$\tilde{\mathbf{w}}\star = \underset{\tilde{w}}{argmax} \underset{n}{min} \left(\frac{t_n \cdot \tilde{\mathbf{w}}' \cdot \tilde{\mathbf{x}}}{||\mathbf{w}||} \right)$ Unfortunately this is a pretty tricky problem to optimize.

Linear SVM

Since signed distance is invariant to the scaling of $\tilde{\mathbf{w}}$, (this is because we divide by its norm), we can choose a scalar λ such that $t_m \cdot (\tilde{\mathbf{w}}' \cdot \tilde{\mathbf{x}}_{\mathbf{m}}) = 1$ for the pair $(t_m, \mathbf{x_m})$ that is closest to the margin! This makes the problem a lot easier to solve since now we must just minimize $min_n \, d_n = \frac{1}{||\mathbf{w}||}$ to maximize the margin. This is a quadratic problem and thus a convex one. It is quadratic since this is equivalent to minimizing $\frac{1}{2} \|\mathbf{w}\|^2$

Hard Margin SVM

This is the case when the data is - linearly separable - there are no outliers

Soft Margin SVM

This is when we add the slack variable ζ. The constraint previously (no outliers) can now be broken and we allow some outliers. **problem:** if we allow ζ to be too big, then we allow too many outliers.

Logistic Regression vs Linear SVM

In logistic regression, training points can come close to the decision boundary. Linear SVM seeks to avoid that.

In practice, maximizing the minimum distance isn't always possible. We must relax some constraints. **We will allow some training points to be misclassified**.

To account for this, we will introduce a **slack variable** ξ_n for each sample. We re-write the constraint as $t_n \cdot (\tilde{\mathbf{w}} \cdot \mathbf{x_n}) \geq 1 - \xi_n$ (instead of just 1 previously.)

We get the following properties:

- if $0 < \xi_i \leq 1$ then sample i lies inside margin, but correctly classified.
- if $\xi_i \geq 1$ then sample i is misclassified.

To be able to correctly train our model, we need to "punish" misclassified points *(otherwise they would occur without consequence)*

Max margin classifiers (*linear SVM*) involve maximizing a convex margin function

$$\underset{w,b}{max} \|w\|^{-2}$$

SVMs are effective on small training sets.

Formulation of the Problem (Linear SVM)

We look for optimal parameters $\mathbf{w}\star$

$$\mathbf{w}^\star = argmin_{(w,\xi_n)} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{n=1}^{N} \xi_n$$

Where C is a constant that controls how costly constraint violations are! We note that this problem is still convex (and easily optimizable) We choose C parameter using cross-validation.

We get a trade off between the number of classification mistakes and the size of the margin.

the points that land on the margin are called the support vectors Only the support vectors determine the decision boundary.