

Lab 5 - Ensembles et SVMs

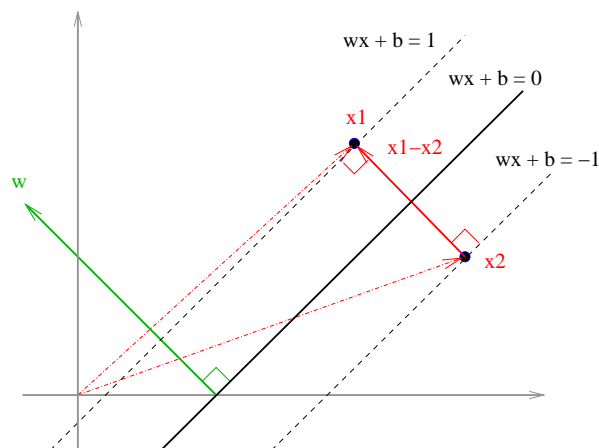
{bengio,mkeller}@idiap.ch
<http://www.idiap.ch/~{bengio,mkeller}>

February 2, 2006

SVMs

1. Show that the margin of a Support Vector Machine is $\frac{2}{\|w\|}$ wide.

.....



Let \mathbf{x}_1 a point on the line with equation:

$$\mathbf{w} \cdot \mathbf{x} + b = 1, \quad (1)$$

and \mathbf{x}_2 the mirror point with respect to the classification hyperplane on the line with equation:

$$\mathbf{w} \cdot \mathbf{x} + b = -1 \quad (2)$$

As illustrated in the figure, the margin $\rho = \|\mathbf{x}_1 - \mathbf{x}_2\|$.

$$(1) - (2)$$

$$\mathbf{w} \cdot \mathbf{x}_1 + b - \mathbf{w} \cdot \mathbf{x}_2 - b = 1 + 1$$

$$\begin{aligned}\mathbf{w} \cdot (\mathbf{x}_1 - \mathbf{x}_2) &= 2 \\ \|\mathbf{w}\| \|\mathbf{x}_1 - \mathbf{x}_2\| \cos(\mathbf{w}, \mathbf{x}_1 - \mathbf{x}_2) &= 2 \\ \rho = \|\mathbf{x}_1 - \mathbf{x}_2\| &= \frac{2}{\|\mathbf{w}\|}\end{aligned}$$

.....

2. How would you control the capacity of an SVM?

Ensembles

3. Implementing bagging.

- (a) Implement the functions defined in `bag.py` in order to make `bagtest.py` work.
- (b) Test with your favorite dataset (prefer a small one to debug).

.....

See `bag_solution.py`.

.....

4. Show how bagging estimator tends to decrease the variance of a regressor learned using mean squared loss function.

(Hint: cf lab1)

.....

Let say that instead of bootstraps, we B_1, \dots, B_b, \dots training sets generated by the same distribution \mathcal{P} . Let $h(X; B)$ be the regressor optimized over B , and

$$h_A(X) = E_B[h(X; B)] = \sum_B h(X; B)$$

the aggregated estimator of the regressor. Over the distribution of any new data point X , the average variance of the predictors $h(X; B)$ is:

$$s^2 = E_B E_X [(h(X; B) - E[h(X; B)])^2] = E_X E_B [(h(X; B) - E_X[h(X; B)])^2].$$

The variance of the aggregated predictor is:

$$s_A^2 = E_X [(h_A(X) - E[h_A(X)])^2] = E_X [(E_B[h(X; B)] - E_X[h(X; B)])^2]$$

We know that $(EZ)^2 \leq E[Z^2]$, thus:

$$s_A^2 \leq s^2$$