SVMs

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

Let \( x_1 \) a point on the line with equation:

\[
wx + b = 1,
\]

and \( x_2 \) the mirror point with respect to the classification hyperplane on the line with equation:

\[
wx + b = -1
\]

As illustrated in the figure, the margin \( \rho = \|x_1 - x_2\| \).

\[
(1) - (2)
\]

\[
w \cdot x_1 + b - w \cdot x_2 - b = 1 + 1
\]
\[ \mathbf{w} \cdot (\mathbf{x}_1 - \mathbf{x}_2) = 2 \]
\[ \|\mathbf{w}\| \|\mathbf{x}_1 - \mathbf{x}_2\| \cos(\mathbf{w}, \mathbf{x}_1 - \mathbf{x}_2) = 2 \]
\[ \rho = \|\mathbf{x}_1 - \mathbf{x}_2\| = \frac{2}{\|\mathbf{w}\|} \]

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 have \(B_1, \ldots, B_b, \ldots\), being \(M\) training sets generated by the same distribution \(\mathcal{P}\). Let \(h(X; B)\) be the regressor optimized over \(B\), and

\[ h_A(X) = E_B[h(X; B)] = \frac{1}{M} \sum_B h(X; B) \]

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:

\[ s^2 = E_B E_X \left[ (h(X; B) - E_X [h(X; B)])^2 \right] = E_X E_B \left[ (h(X; B) - E_X [h(X; B)])^2 \right] \]

The variance of the aggregated predictor is:

\[ s_A^2 = E_X \left[ (h_A(X) - E_X [h_A(X)])^2 \right] = E_X \left[ (E_B \{h(X; B) - E_X [h(X; B)]\})^2 \right] \]

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

\[ s_A^2 \leq s^2 \]