Machine Learning
Assignment 3.2

Nikolaj Dybdahl Rathcke (rfq695)

January 14, 2016

1 Margin-based VC bound for SVMs

Theorem 2.6 in the lecture provides the following VC generalization bound for SVM's:

$$\mathbb{P}\left\{\exists h \in \mathcal{H} : L(h) \geq \hat{L}(h,S) + \sqrt{\frac{8\ln(2\left((2n)^{d_{VC}}+1\right)/\delta)}{n}}\right\} \leq \delta$$

When our input space $\mathcal{X}$ is in a ball of radius $R = 1$, the lecture notes has derived the bound $d_{VC} = i+1$, where i is the i'th subspace in our nested sequence of subspaces.

Theorem 2.7 states that:

$$d_{VC}(\mathcal{H}_\rho) \leq \lceil R^2/\rho^2 \rceil + 1$$

The hypothesis space $\mathcal{H}_i$ has a margin which is larger than $\frac{1}{\rho^2} = i$ for $R = 1$. We can use theorem 2.7 and the same lower bound on the margin to get $d_{VC}(\mathcal{H}_i) = R^2 i + 1$. Exchanging this in theorem 2.6 yields:

$$\mathbb{P}\left\{\exists h \in \mathcal{H} : L(h) \geq \hat{L}(h,S) + \sqrt{\frac{8\ln(2\left((2n)^{R^2 i+1}+1\right)/\delta)}{n}}\right\} \leq \delta$$

Which is a bound for a general R.

2 Occam's razor bound and the lower VC bound

2.1

The VC-dimension is 27^d as $\mathcal{H}_d$ is a tree with depth d and branches 27 times at every node. Since we're only looking at words of length d, the VC-dimension is equal to the number of leaves, i.e. 27^d.

2.2

The VC-dimension is ∞. If we construct a tree as in (2.1), the tree would have infinite depth as we look at infinite d.

2.3

The idea of Occam's razor bound is to give each hypothesis a probability. This probability is given by $p(h)$ where $\sum_{h \in \mathcal{H}} p(h) \leq 1$ and as such, the "burden" is shared by all hypotheses with a specific weight.

3 Neural Networks

3.1 Neural network implementation

An attempt at an implementation has been made in `src/nn.py`, which does not provide the right answer.

3.2 Neural network training

N/A