Lecture 5

Huang Ziheng

November 7, 2017

1 AdaBoost

Adaptive Boosting
Multiplicative Weight Updating, Greedy optimize exponential loss

$$\text{Exponential loss} = \frac{1}{n}\sum_{i=1}^{n} exp\left\{-y_i f(x_i)\right\}$$
$$f(x) = \sum_{t=1}^{T}\alpha_t h_t(x) \text{ And } F(x) = sign[f(x)]$$

For $F \in CH_\tau(\mathcal{H})$, Obviously VC dimension grows at leat linearly by the sample size, $VC[CH_\tau(\mathcal{H})] > n$ so when the sample size is large the VC-dimension is too large.
And by practice, even when the training error is 0 when we continue to add samples the test error will still goes down, no overfitting problems.

2 Margin Theory for Boosting(Voting Classifiers)

Function:

$$yf(x) = y\sum_t \alpha_t h_t(x)$$

is another distance but not Euclid of course.

$$x \mapsto (h_1(x), ...h_T(x)), h_t(x) = \pm 1$$
$$yf(x) \text{ is a distance for } (x, y) \text{ defined by } (\alpha_1, ..., \alpha_T)$$

Homework1: Distance in Boosting

If for most (x_i, y_i), $y_i f(x_i)$ is large then f has good generalization ability, which is called data dependent generalization:
For $CH(\mathcal{H})$, key idea: Approximation, to find a set that has a small VC-Dimension and is close to convex τ, that is: $CH_N(\mathcal{H}) \approx CH(\mathcal{H})$
Notice that: $CH(\mathcal{H}) = \sum \alpha_t h_t, \alpha_t \geq 0, \sum \alpha_t = 1$
Then we have: $CH_N(\mathcal{H}) = \frac{1}{N}\sum_{i=1}^{N} h_{t_i}, h_{t_i} \in \mathcal{H}$
For given x, $\forall f \in CH(\mathcal{H}), \exists g \in CH_N(\mathcal{H}), f(x) \approx g(x)$

Thm: Assume $|\mathcal{H}| < \infty$, then with probability $1-d$ over the following inequalities holds simultaneously for all $f \in CH(\mathcal{H})$ and all $\theta \in (0,1]$
$P_D(yf(x) \leq 0) \leq P_S(yf(x) \leq 0) + O(\frac{1}{\sqrt{n}}(\frac{log(n)log|\mathcal{H}|}{\theta^2} + log\frac{1}{\delta})^2)$
Which means that if for most data the margin is small then this classifier is good.

To prove, steps:
①$f \in CH(\mathcal{H}), w.h.p.\ g \in CH_N(\mathcal{H}), f(x) \approx g(x)$
So we want to prove that:
$P_D(yf(x) \leq 0) \leq P_D(yg(x) \leq \frac{\theta}{2}) + small$
②$P_D(yg(x) \leq \frac{\theta}{2}) \leq P_S(yg(x) \leq \frac{\theta}{2})$+ Complexity of $CH_N(\mathcal{H})$
③$P_S(yg(x) \leq \frac{\theta}{2}) \leq P_S(yf(x) \leq \theta) + small$
Rob Schapire in Microsoft, 1990, The strength of Weak Learnable. Take the prove in xitike

3 Term project 2

Fundamental background: Neural network in our brain is not too deep and calculation efficiency is far better, and is highlight distributed instead of central control system. So I want to compare human brain with computer, although we may reach the complexity of brain but our efficiency and control system is still not comparative to brain. So there are some advance system or algorithm in brain we can explore.

We have many concepts in our mind, whether concrete or abstract, and how these concepts performs is still a mystery. Then how to put concepts into our algorithm is an interesting question. Now deep learning performs well because it is quite the same way human brain used to deal with imagines. But to go further we need to explore concepts.(Les Valiant, The Circuit of the Mind).Concepts are connected, and can be classified into deep concepts and surface concepts and so on.

Hinton, Capsule. Reference papers:
① Dynamic Routing Between Capsules. NIPS'17.
② Matrix Capsules with EM Routing, ICLR'18 submission.

Our work is to assume that concepts is formed by the connection by neural networks, and then explore how to put concepts into our algorithm. But our methods are not limited on these two papers, so we can design new algorithms for learning network with capsules. Furthermore there work are focused on simple training data, so we can find experimental results at bench mark datasets.

4 (

PAC-Bayes Theory) Bayesion vs. Frequentist
Prior distribution and posterior distribution
Let Q be a distribution of classifiers
$textcircled1$ Stochastic classifier, $error_D(f_Q)$
$textcircled2$ Voting classifier, $error_D(v_Q)$
Homework 2: find the relationship between these two errors.