IT585 Advanced Machine Learning

Classes : Mon, Thu, Fri 9 AM – 9:50 AM

Instructor : Rachit Chhaya

FB – 3 3109

Reference Books:

1) Machine Learning : A probabilistic perspective by Kevin Murphy
2) Learning from Data: A short course by Abu Mostafa, Ismail and Lin
3) Understanding Machine Learning by Shai Shalev Shwartz and Ben David
4) Pattern Recognition and Machine Learning by Bishop

Topics (Broadly)

- Learning Theory
- Optimization for ML (SVM Revisited)
- GMMs, MLE, MAP, EM
- Ensemble Techniques
- ANNs
- Spectral Clustering

— More Advanced Topics

This course requires familiarity with mathematics. We will frequently use it.

— We will not talk much about Deep Networks (separate course)

Evaluation Scheme

- Term paper/Project — 20%
- Reproducibility Challenge — 15%
- Lab/Homework Assignments — 10%
- Quizzes — 20%
- Scribe Notes — 10%
- End Sem — 25%

A Statistical Learning Framework

- Domain Set - X
- Label set - Y
- Training Data: $S = ((x_1, y_1), (x_2, y_2) \ldots (x_m, y_m))$ from $X \times Y$
- Learner's Output: $h : X \to Y$

- Data Generation Model:

  Prob. Dist. over X is D

  Labeling function $f: X \to Y$

- Measures of Success

$$L_{D,f}(h) = \underset{x \sim D}{P}\left[h(x) \neq f(x)\right]$$

ERM Framework

Training Error

$$L_S(h) = \frac{|\{i \in [m] : h(x_i) \neq y_i\}|}{m}$$

Search for a solution that works well on available data

Hypothesis Class ($\mathcal{H}$)

- Restricted search space

I I D Assumption

- Independently and Identically Distributed

Agnostic PAC Learning

Probably Approximately correct

$\mathcal{H}$ is PAC Learnable if

- $m_{\mathcal{H}} : (0,1)^2 \to \mathbb{N}$
- $\epsilon, \delta \in (0,1)$, D over $X \times Y$

When running the learning algorithm on $m \geq m_{\mathcal{H}}(\epsilon, \delta)$, algorithm returns h s.t.

$$L_{\mathcal{D}}(h) \leq \min_{h' \in \mathcal{H}} L_{\mathcal{D}}(h') + \epsilon$$

ϵ - representative Sample

Training set S is called ϵ - representative if

$$\forall h \in \mathcal{H} \quad |L_S(h) - L_D(h)| \leq \epsilon$$

Finite Classes are Agnostic PAC Learnable

Proof: Requires

i) Union Bound

ii) Hoeffdings Inequality

Union Bound

For any two sets A, B and a distribution D we have

$$D(A \cup B) \leq D(A) + D(B)$$

Hoeffding's Inequality

Let $\theta_1, \ldots \theta_m$ be a seq. of i.i.d random variables and assume that $\forall i, E[\theta_i] = \mu$ and $P[a \leq \theta_i \leq b] = 1$. Then for any $\epsilon > 0$

$$P\left[\left|\frac{1}{m}\sum_{i=1}^{m}\theta_i - \mu\right| > \epsilon\right] \leq 2\exp\left(\frac{-2m\epsilon^2}{(b-a)^2}\right)$$


Proof !!