机器学习基石笔记

Jacky Wang

March 4, 2016

Contents

When Can Machines Learn?

The Learning Problem

Components of Machine Learning

$\mathcal{A}$ takes $\mathcal{D}$ and $\mathcal{H}$ get $g \approx f$

- ***A***lgorithm $\mathcal{A}$
- ***D***ata $\Longleftrightarrow$ Training examples: $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N)\}$
- ***H***ypothesis set $\mathcal{H}$
- final hypothesis g, where
  - $g \in \mathcal{H}$
  - $g \approx f$
  - $g : \mathcal{X} \to \mathcal{Y}$
- unknown target ***f***unction: $f : \mathcal{X} \to \mathcal{Y}$

Perceptron Learning Algorithm

- $\mathcal{H} = \{\text{sign}(h(\mathbf{x}; \mathbf{w}))\}$ where:
  - $\mathbf{w} \in \mathcal{R}^{d+1}$
  - $\mathbf{x} = (1, x_1, x_2, ..., x_d)^T := (x_0, x_1, x_2, ..., x_d)^T$
- $\dagger \in \{+1, -1\}$

$$\frac{\mathbf{w}_f^T \mathbf{w}_T}{||\mathbf{w}_f|| ||\mathbf{w}_T||} = cos(\theta_{f,T}) \tag{1}$$

$$\mathbf{w}_T = \mathbf{w}_{T-1} + y_{n(T)}\mathbf{x}_{n(T)} = \mathbf{w}_{T-2} + y_{n(T-1)}\mathbf{x}_{n(T-1)} + y_{n(T)}\mathbf{x}_{n(T)} = \sum_{t=0}^{T} y_{n(t)}\mathbf{x}_{n(t)}$$

$||\mathrm{w}_T||^2 = (2)$

- The rule somewhat "tries to correct the mistake"
- Mistake limits the growth of $||\mathbf{w}_t||^2$, even updating with the longest $\mathbf{x}_n$
-

Learning to Answer Yes/No

Types of Learning

Feasibility of Learning

Why Can Machines Learn? [為什麼機器可以學習]

Training versus Testing [訓練與測試]

Theory of Generalization [舉一反三的一般化理論]

The VC Dimension [VC 維度]

Noise and Error [雜訊一錯誤]

How Can Machines Learn? [機器可以怎麼樣學習]

Linear Regression [線性迴歸]

Linear 'Soft' Classification [軟性的線性分類]

Linear Classification beyond Yes/No [二元分類以外的分類問題]

Nonlinear Transformation [非線性轉換]

How Can Machines Learn Better? [機器可以怎麼樣學得更好]

Hazard of Overfitting [過度訓練的危險]

Preventing Overfitting I: Regularization [避免過度訓練一：控制調適]

Preventing Overfitting II: Validation [避免過度訓練二：自我檢測]

Three Learning Principles [三個機器學習的重要原則]