Parametric Models: from data to models

Pradeep Ravikumar (Instructor), HMW-Alexander (Noter)

January 23, 2017

Back to Index

Contents

Resources

- Lecture

---

1 Recall Model-based ML

Model-based ML

2 Model Learning: Data to Model

Questiongs:

- What are the principles in going from data to model?
- What are the guarantees of these methods?

2.1 Bernoulli Distribution Example

- Bernoulli distribution model
  - X is a random variable with Bernoulli distribution when:
    - X takes values in $\{0, 1\}$
    - $P(X = 1) = \theta,\ P(X = 0) = 1 - \theta$
    - Where $\theta \in [0, 1]$

- Draw **independent** samples that are **identically distributed** from same distribution model, Bernoulli distribution.
  - If we observe an event $X \in \{0, 1\}$, its probability $P(X)$ is $\theta^X(1-\theta)^{1-X}$
  - Then the probability of data:

$$\begin{aligned} \mathbb{P}(X_1, X_2, ..., X_n; \theta) &= \prod_{i=1}^{n} P(X_i) \\ &= \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} \\ &= p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i} \\ &= p^{n_1}(1-p)^{n-n_1} \end{aligned} \tag{1}$$

- Maximum Likelihood ($p(D|\theta)$) Estimator (MLE)
  - Choose θ that maximizes the probability of observed data.

$$\begin{aligned} \hat{\theta} &= \arg\max_\theta \mathbb{P}(X_1, \ldots, X_n; \theta) \\ &= \arg\max_\theta \theta^{n_1}(1-\theta)^{n-n_1} \\ &= \arg\max_\theta n_1 \log\theta + (n-n_1)\log(1-\theta) \\ &\Rightarrow \frac{n_1}{\hat{\theta}} - \frac{n-n_1}{1-\hat{\theta}} = 0 \\ &\Rightarrow \hat{\theta}_{MLE} = \frac{n_1}{n} \end{aligned} \tag{2}$$

2.2 How good is this MLE?

- Consistency:
  - As we sample more and more times, we want our estimator to converge (in probability) to the true probability.
  - For Bernoulli distribution example, we get the $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \theta$ in probability as $n \to \infty$ by the **Law of Large Numbers**!

3 How Good is the Maximum Likelihood Estimation?

3.1 Unbiasedness

3.2 title