5.6: Bayesian Statistics

Sunday, June 10, 2018 10:09 PM

until now, frequentist statistics - make all predictions based on a single value of θ.

Bayesian statistics - consider all possible values of θ to make prediction

prior probability distribution, $p(\theta)$ "prior"

what we know before seeing data

e.g., uniform distribution or Gaussian with high entropy

observe data $\{x^{(1)}, \dots, x^{(m)}\}$

update belief about θ using Bayes' rule:

$$p(\theta \mid x^{(1)}, \dots, x^{(m)}) = \frac{p(x^{(1)}, \dots, x^{(m)} \mid \theta)\, p(\theta)}{p(x^{(1)}, \dots, x^{(m)})}$$

Comparison to maximum likelihood estimation:

1. MLE makes predictions with a point estimate of θ, Bayesian estimation makes predictions using full distribution over θ
   incorporates uncertainty by integrating over possibilities
2. contribution of prior distribution
   shifts probabilities to ones preferred a priori

Typically, Bayesian generalizes better with limited data, but high cost

Bayesian Linear Regression

(5.6.1) Maximum A Posteriori (MAP) Estimation

Bayesian → choose single point that is maximum (posterior) probability:

$$\theta = \arg\max p(\theta / x) = \arg\max \log p(x \mid \theta) + \log p(\theta).$$

$$\theta_{MAP} = \arg\max_{\theta} p(\theta | x) = \arg\max_{\theta} \underbrace{\log p(x | \theta)}_{\text{log likelihood}} + \underbrace{\log p(\theta)}_{\text{prior}}.$$