

Maximum Likelihood vs. Bayesian

- Maximum likelihood estimation

- “Best” means “data likelihood reaches maximum”

$$\hat{\theta} = \arg \max_{\theta} P(\mathbf{X} | \theta)$$

- Problem: Small sample

- Bayesian estimation:

Bayes Rule

$$p(\mathbf{X} | \mathbf{Y}) = \frac{p(\mathbf{Y} | \mathbf{X})p(\mathbf{X})}{p(\mathbf{Y})}$$

- “Best” means being consistent with our “prior” knowledge and explaining data well

$$\hat{\theta} = \arg \max_{\theta} P(\theta | \mathbf{X}) = \arg \max_{\theta} P(\mathbf{X} | \theta)P(\theta)$$

- Problem: How to define prior?



Maximum a Posteriori (MAP) estimate