Bayesian Inference

Qiang Liu
UT Austin

Bayesian Inference

parameter

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)} \propto p(D \mid \theta)p(\theta).$$

Data.

$P(\theta)$: Prior

Example: Predicting the Commute Time

- You move to a new apartment.
- Your friend told you the commute time is 30 ± 10 min. (Prior) ☆
- You also drove yourself a few time, and found the time is 25, 45, 30, 50, ... ☆
- How should you predict the commute time?

θ: time.

Prior: $P(\theta) \sim \mathcal{N}(\mu_0, \sigma_0^2)$

$$\boxed{\mu_0 = 30 \quad \sigma_0 = 10}$$

observe: $\{x_1 \cdots x_n\} \triangleq D$

$x_i = \theta + \sigma_1 \xi_i \qquad \xi_i \sim \mathcal{N}(0, 1) \qquad \boxed{\sigma_1 = 5}$

$P(x_i|\theta) \sim \mathcal{N}(\theta, \sigma_1^2)$

Posterior:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \propto P(D|\theta)P(\theta)$$

$$= \underbrace{\left[\prod_{i=1}^n P(x_i|\theta)\right]}_{P(D|\theta)} P(\theta)$$

$\mathcal{N}(\theta, \sigma_1^2)$

$$P(x_i|\theta) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x_i-\theta)^2}{2\sigma_1^2}\right)$$

$$\propto \exp\left(-\frac{(x_i-\theta)^2}{2\sigma_1^2}\right)$$

$$\propto \left[\prod_{i=1}^n \exp\left(-\frac{(x_i-\theta)^2}{2\sigma_1^2}\right)\right]\exp\left(-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right)$$

$\theta^2 - 2x_i\theta + x_i^2$

$$\propto \exp\left(-\sum_{i=1}^n \frac{(\theta-x_i)^2}{2\sigma_1^2} - \frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right)$$

$$= \exp\left(-\frac{1}{2}(A\theta^2 - 2B\theta + C)\right)$$

$$= \exp\left(-\frac{1}{2}A\left(\theta - \frac{B}{A}\right)^2 + \text{const}\right)$$

$$\sim \mathcal{N}\left(\frac{B}{A}, \frac{1}{A}\right)$$

$$A = \sum_{i=1}^n \frac{1}{\sigma_1^2} + \frac{1}{\sigma_0^2} = \frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}$$

$$B = \sum_{i=1}^n \frac{x_i}{\sigma_1^2} + \frac{\mu_0}{\sigma_0^2}$$

$C = ??$

$$\boxed{\mu_p = \frac{B}{A} = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_1^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}} \qquad \sigma_p^2 = \frac{1}{A} = \left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}\right)^{-1}}$$

If $n = 0$ (No data), $\mu_p = \frac{\mu_0}{\sigma_0^2} \Big/ \frac{1}{\sigma_0^2} = \mu_0$.

$\sigma_p^2 = \sigma_0^2$

If $n > 0$ (have data)

If $n \to +\infty$ (Infinite data) $\mu_p \approx \frac{\sum_{i=1}^n \frac{x_i}{\sigma_1^2} + 0}{\frac{n}{\sigma_1^2} + 0} \approx \frac{\sum_{i=1}^n x_i}{n}$.

$\sigma_p^2 = \left(\frac{n}{\sigma_1^2}\right)^{-1} = \frac{\sigma_1^2}{n}$

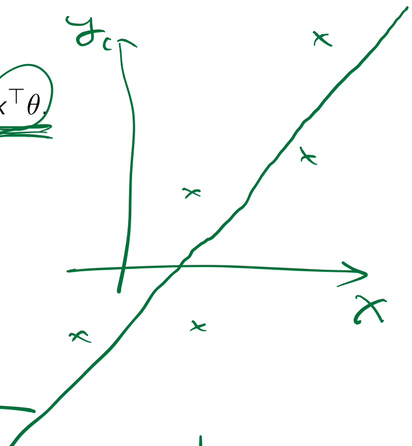
$$_p \sim \frac{}{\sigma_i^2} = \frac{}{n}$$

Bayesian Linear Regression

- Given data points $\underbrace{\{x_i, y_i\}_{i=1}^n}_{\triangleq D}$, want to find θ, such that $y \approx x^\top\theta$.

Least Square:

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} (y_i - x_i^T\theta)^2$$



$$y = x\theta + b = \underbrace{[x, 1]^T}_{\tilde{x}} \underbrace{\begin{bmatrix}\theta \\ b\end{bmatrix}}_{\tilde{\theta}} = \tilde{x}^T\tilde{\theta}$$

Bayesian Inference:

Treat θ as a Random Variable.

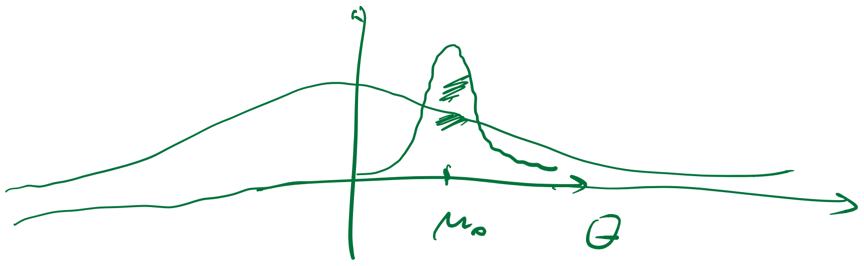
Assume prior:

$$P(\theta) \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$\boxed{\mu_0 = 0. \quad \sigma_0^2 = \text{Large Number.}}$

Assume Likelihood:

$$y_i = x_i^T\theta + \sigma_1 \xi_i$$



σ_1 : Variance

$\xi_i \sim \mathcal{N}(0, 1)$

$$P(\{y_i, x_i\} \mid \theta) = P(y_i \mid x_i, \theta)\, P(x_i)$$

$$P(\theta \mid D) \propto P(D \mid \theta)\, P(\theta)$$

$$= \left[\prod_{i=1}^{n} P(\{y_i, x_i\} \mid \theta)\right] P(\theta)$$

$$= \left[\prod_{i=1}^{n} P(y_i \mid x_i, \theta)\, P(x_i)\right] P(\theta)$$

$$\propto \left[\prod_{i=1}^{n} P(y_i \mid x_i, \theta)\right] P(\theta)$$

$$\propto \left[\prod_{i=1}^{n} \exp\left(-\frac{(y_i - x_i^T\theta)^2}{2\sigma_1^2}\right)\right] \exp\left(-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right)$$

$$= \exp\left(-\sum_{i=1}^{n} \frac{(y_i - x_i^T\theta)^2}{2\sigma_1^2} - \frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right)$$

$\theta = \begin{bmatrix} \\ \\ \end{bmatrix}$ $d \times 1$

$$= \exp\left(-\frac{1}{2}\left(\theta^T A \theta - 2B^T\theta + \text{const}\right)\right)$$

$$\sim \mathcal{N}(A^{-1}B,\ A^{-1})$$

$$A = \sum_{i=1}^{n} \frac{x_i x_i^T}{\sigma_1^2} + \frac{I}{\sigma_0^2}$$

$$B = \sum_{i=1}^{n} \frac{y_i x_i}{\sigma_1^2} + \frac{\mu_0}{\sigma_0^2}$$

$A = \begin{bmatrix} & & \end{bmatrix}$ $d \times d$

$$\mu_p = A^{-1}B$$

$$\sigma_p^2 = A^{-1}$$