Bayesian Models

A Quick Review...

Frequentist Approach $\underset{\theta}{\operatorname{argmax}}\ P(X|\theta)$

Bayesian Approach $\underset{\theta}{\operatorname{argmax}}\ \underbrace{P(\theta|X)}_{\text{posterior}}$

In Bayesian models, we take the Bayesian approach to estimating parameters.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ P(\theta|X) = \underset{\theta}{\operatorname{argmax}}\ \frac{P(X|\theta)P(\theta)}{P(X)}$$

$$= \underset{\theta}{\operatorname{argmax}}\ \underbrace{P(X|\theta)}_{\text{likelihood}}\underbrace{P(\theta)}_{\text{prior}}$$

Posterior Mean

$E[\theta|X]$

Reason for using a prior: computational convergence

$$p(\theta|X) = \frac{P(X|\theta)P(\theta)}{p(X)\ \left\{ \int P(X|\theta)P(\theta)d\theta \right.}$$

If we don't assume a distribution over θs, we must integrate over all possible prior distributions.

If we choose a prior, the integral in the denominator becomes tractable

$$\therefore\ P(\theta|X) \propto P(X|\theta)P(\theta)$$

How do I choose a prior?

Typically, the posterior and the prior are from the same family of distributions.

A prior is a conjugate prior for a given likelihood distribution if the resulting posterior has the same form as the prior.

Example: Poisson-Gamma Conjugacy

Poisson: Likelihood $\quad P(X|\theta) = \dfrac{e^{-n\lambda}\lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!} = P(X|\lambda)$

Gamma: Prior $\quad P(\theta) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} = P(\lambda|\alpha,\beta)$

$$P(X|\theta)P(\theta) = \left(\frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}\right)\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\right)$$

$$= \frac{\beta^{\alpha}\lambda^{\Sigma x_i+\alpha-1}e^{-\lambda(n+\beta)}}{\prod_{i=1}^{n} x_i!\,\Gamma(\alpha)} \underset{\substack{\uparrow \\ \text{proportional} \\ \text{to}}}{\propto} \frac{(n+\beta)^{\alpha}\lambda^{\Sigma x_i+\alpha-1}e^{-\lambda(n+\beta)}}{\Gamma(\Sigma x_i+\alpha)}$$

$$= \text{Gamma}(\Sigma x_i+\alpha,\ n+\beta) \rightarrow P(\theta|X)$$