Bayesian Machine Learning

1. Model specification.

model $p(x, \theta \mid m)$

$$\begin{cases} p(x \mid \theta, m) & \text{data generating distr.} \\ p(\theta \mid m) & \text{parameter prior distr.} \end{cases}$$

2. parameter estimation

$D = \{x_1, \ldots, x_N\}$

$$p(\theta \mid D, m) = \frac{p(D \mid \theta, m)\, p(\theta \mid m)}{p(D \mid m)}$$

$$p(D \mid \theta, m) = \prod_{n=1}^{N} p(x_n \mid \theta, m)$$

$$p(D \mid m) = \int p(D, \theta \mid m)\, d\theta$$

$$= \int \underbrace{p(D \mid \theta, m)\, p(\theta \mid m)}_{\text{numerator}}\, d\theta$$

3. model evaluation

$\mathcal{M} = \{m_1, m_2, \ldots, m_K\}$

$p(m_k \mid D)$

model spec: $p(x \mid \theta, m_k)$, $p(\theta \mid m_k)$, $p(m_k)$

model prior

$$p(m_k \mid D) = \frac{p(D \mid m_k)\, p(m_k)}{p(D)}$$

$$\propto p(D \mid m_k)\, p(m_k)$$

$$= \int p(D, \theta \mid m_k)\, d\theta \cdot p(m_k)$$

$$= p(m_k) \int p(D \mid \theta, m_k)\, p(\theta \mid m_k)\, d\theta$$

4. application (prediction)

$$p(x_{N+1} \mid D) =$$
$$= \int p(x_{N+1}, \theta \mid D)\, d\theta$$
$$= \int \underbrace{p(x_{N+1} \mid \theta)}_{\text{data gen. distr.}} \cdot \underbrace{p(\theta \mid D)}_{\text{par. posterior}}\, d\theta$$

ML is <u>easy</u>!, apart from computational details.