Probabilistic & Unsupervised Learning

October 2, 2018

1 Lecture 1

1.1 A probabilistic approach

$P(x|\theta)$ is the generative model.

$P(y|x,\theta)$ is the likelihood.

1.2 Bayesian learning

Data: $D = x_1, \ldots, x_n$

Model: M_1, M_2, etc

Parameters: θ_i (per model)

Prior probability of models: $P(M_i)$

Prior Probability of model parameters: $P(\theta_i|M_i)$

Model of data given parameters (likelihood model): $P(x|\theta_i, M_i)$

Data probability (likelihood)

$$P(D|\theta_i, M_i) = \prod_{j=1}^{n} P(x_j|\theta, M_i) \equiv \iota(\theta_i)$$

Parameter learning (posterior (based on the condition))

$$P(\theta|D, M_i) = \frac{P(D|\theta_i, M_i)P(\theta|M_i)}{P(D|M_i)}$$

$$P(D|M_i) = \int P(D, \theta_i|M_i)d\theta_i = \int P(D|\theta_i, M_i)P(\theta_i|M_i)d\theta_i$$

$P(D|M_i)$ is called the marginal likelihood or evidence for M_i. In the second formula, it needs to transform into the form of joint distribution function integration at first.

model selection

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)}$$

Beta distribution

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$$

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

Conjugate priors (why it is easy to convert the posterior)

Definition: In Bayesian probability theory, if the posterior distributions $p(\theta|x)$ are in the same probability distribution family as the prior probability distribution $p(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function

exponential family likelihood (including Binomial, Normal, Gamma distribution)

$$P(x|\theta) = g(\theta)f(x)e^{\phi(\theta)^{\mathrm{T}}\mathbf{T}(x)}$$

where $g(\theta)$ is the normalizing constant.

$$P(\{x_i\}|\theta) = \prod_i P(x_i|\theta) = g(\theta)^n e^{\phi(\theta)^{\mathrm{T}}(\sum_i \mathbf{T}(x_i))} \prod_i f(x_i)$$

If the prior takes the conjugate form.

$$P(\theta) = F(\tau, \nu)g(\theta)^\nu e^{\phi(\theta)^{\mathrm{T}}\tau}$$

with $F(\tau, \nu)$ the normalizer, then posterior is

$$P(\theta|\{x_i\}) \propto P(\{x_i\}|\theta)P(\theta) \propto g(\theta)^{\nu+n}e^{\phi(\theta)^{\mathrm{T}}(\tau+\sum_i \mathbf{T}(x_i))}$$

where $F(\tau + \sum_i \mathbf{T}(x_i), \nu + n)$ is the normalizer.