Reading Notes for ch2 Probability Distributions

Xiang Pan

September 10, 2021

1 Priors

1.1 Parametric Priors

Conjugate priors: lead to posterior distributions having the same functional form as the prior.

- Multinomial distribution: Dirichlet distribution
- Gaussian distribution: Gaussian distribution

For the Exponential Family, conjugate prior,

$$p(\boldsymbol{\eta} \mid \boldsymbol{\chi}, \nu) = f(\boldsymbol{\chi}, \nu) g(\boldsymbol{\eta})^{\nu} \exp\left\{\nu \boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{\chi}\right\}$$

conjugate posterior,

$$p(\boldsymbol{\eta} \mid \mathbf{X}, \boldsymbol{\chi}, \nu) \propto g(\boldsymbol{\eta})^{\nu+N} \exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\left(\sum_{n=1}^{N} \mathbf{u}\left(\mathbf{x}_{n}\right)+\nu \boldsymbol{\chi}\right)\right\}$$

For the parameter estimation, posterior mean of θ, averaged over the distribution generating the data, is equal to the prior mean of θ.

$$\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta}] = \mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \mid \mathcal{D}]\right]$$

$$\operatorname{var}_{\boldsymbol{\theta}}[\boldsymbol{\theta}] = \mathbb{E}_{\mathcal{D}}\left[\operatorname{var}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \mid \mathcal{D}]\right] + \operatorname{var}_{\mathcal{D}}\left[\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \mid \mathcal{D}]\right]$$

1.2 Nonparametric Density Estimation

Notation Rigion R, total number K of points that lie inside R. V is the volume of R, N is the dataset number.

We can decide the K and V from the data by fixing V and determining the value of K from the data(kernel density estimator) or fixed value of K and use the data to find an appropriate value for V(Nearest-neighbour methods).

kernel density estimator(e.g. Gaussian kernel density estimator): Gaussian density model is obtained by placing a Gaussian over each data point and then adding up the contributions over the whole data set, and then dividing by N.

Nearest-Neighbour methods [1]

2 Gaussian Distribution

Multivariate Gaussian distribution is that if two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is again Gaussian. Similarly, the marginal distribution of either set is also Gaussian.

[1] Limited to the page limit, you can get the full version note at https://github.com/Xiang-Pan/NYU_Baysian_Machine_Learning/blob/master/reading_notes/ch2/note2.pdf

2.1 Conditional Gaussian

$$p\left(\mathbf{x}_a \mid \mathbf{x}_b\right) = \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1}\right)$$
$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}\left(\mathbf{x}_b - \boldsymbol{\mu}_b\right)$$

2.2 Marginal Gaussian

$$p\left(\mathbf{x}_a\right) = \mathcal{N}\left(\mathbf{x}_a \mid \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa}\right)$$

2.3 Mixtures of Gaussians

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

$$\ln p(\mathbf{X} \mid \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \right\}$$

The maximum likelihood solution for the parameters no longer has a closed-form analytical solution.

3 Continus Estimation

Robbins-Monro procedure for parameter estimation:

$$\theta^{(N)} = \theta^{(N-1)} + a_{N-1} \frac{\partial}{\partial \theta^{(N-1)}} \ln p\left(x_N \mid \theta^{(N-1)}\right)$$