Machine Learning I Lecture III: Gaussian models

AMIT DHOMNE

Plan for today

Wrap up: Continuous random variables

Gaussians

Bayesian inference for Gaussians

Mean, variance, and conditioning on events are the same as the discrete case, just with sums replaced by integrals.

- Mean: $E(X) = \int_x x \cdot p(x) dx$
- Variance: $\mathrm{Var}(X) = E(X^2) - E(X)^2$
- Example: Uniform, Exponential [on board]
- If X has pdf $p(x)$, then $X|(X \in A)$ has pdf

$$p_{X|A}(x) = \frac{p(x)}{P(A)} = \frac{p(x)}{\int_{x \in A} p(x) dx} \tag{1}$$

- Only makes sense if $P(A) > 0$!
- Examples: Uniform, Exponential [on board]

Bivariate continuous distributions: Marginalization, Conditioning and Independence

- $p_{X,Y}(x, y)$, joint probablity density function of X and Y
- $\int_x \int_y p(x, y)dxdy = 1$
- Marginal distribution: $p(x) = \int_{-\infty}^{\infty} p(x, y)dy$
- Conditional distribution: $p(x|y) = \frac{p(x,y)}{p(y)}$
- Note: $P(Y = y) = 0$! Formally, conditional probability in the continuous case can be derived using infinitesimal events.
- Independence: X and Y are independent if $p_{X,Y}(x, y) = p_X(x)p_Y(y)$

The univariate Gaussian

$$t \sim \mathcal{N}(\mu, \sigma^2) \tag{2}$$

$$p(t|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) \tag{3}$$

- The Gaussian has mean μ and variance σ^2 and precision $\beta = 1/\sigma^2$
- Q: What are the mode and the median of the Gaussian?
- Maximum Likelihood estimation of μ and β: [on board]
- Q: How would you find the conjugate prior for the Gaussian?

A (very important) aside: Products of Gaussian pdfs are (unnormalized) Gaussians pdfs

- Suppose $p_1(x) = \mathcal{N}(x, \mu_1, 1/\beta_1)$ and $p_2(x) = \mathcal{N}(x, \mu_2, 1/1\beta_2)$, then

$$p_1(x)p_2(x) \propto \mathcal{N}(x, \mu, 1/\beta) \tag{4}$$

$$\beta = \beta_1 + \beta_2 \tag{5}$$

$$\mu = \frac{1}{\beta}(\beta_1\mu_1 + \beta_2\mu_2) \tag{6}$$

In general:

$$p_1(x)p_2(x)...p_n(x) \propto \mathcal{N}(x, \mu, 1/\beta) \tag{7}$$

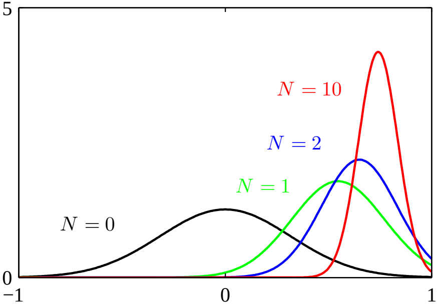
$$\beta = \sum_n \beta_n \tag{8}$$

$$\mu = \frac{1}{\beta}\sum_n \mu_n\beta_n \tag{9}$$

This is also true for multivariate Gaussians!

Bayesian Inference for the Gaussian

- Suppose we are given data $D = \{x_1, \ldots, x_N\}$.
- We assume that the data is Gaussian-distribution with known variance σ^2 and unknown mean μ.
- Our prior for μ is Gaussian: $\mu \sim \mathcal{N}(\mu_o, \sigma_o^2)$
- Posterior distribution over μ given the data: [on board]



[Bishop PRML Figure 2.12]

Behaviour for large N: [on board]

What if the variance is not given?

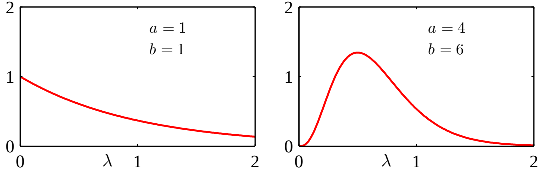
- For simplicity, assume mean to be known.
- More convenient to work with precision $\lambda = 1/\sigma^2$.
- Conjugate prior: Gamma distribution $\mathrm{Gam}(\lambda|a, b)$

$$p(\lambda|a,b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} exp(-b\lambda) \tag{10}$$

- Posterior is $\mathrm{Gam}(\lambda|a_N, b_B)$

$$a_N = a + \frac{N}{2} \tag{11}$$

$$b_N = b + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2 \tag{12}$$

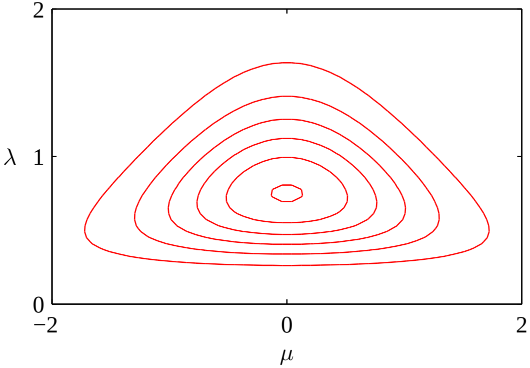


[Bishop PRML Page 100]

What if both the mean and the variance are unknown?

- Conjugate prior: Gaussian-Gamma distribution

$$p(\mu, \lambda) = \mathcal{N}\left(\mu | \mu_o (\beta\lambda)^{-1}\right) \mathrm{Gam}(\lambda | a, b) \tag{13}$$



S
[Bishop PRML Page 102]