Machine Vision

October 14, 2018

1 Lecture 1

Bernoulli Distribution

$$Pr(x) = \lambda^x (1-\lambda)^{1-x}, \lambda \in= [0,1], x \in \{0,1\}$$

$$Pr(x) = Bern_x[\lambda]$$

Beta Distribution

$$Pr(\lambda) = \frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]} \lambda^{\alpha-1}(1-\lambda)^{\beta-1}, \alpha, \beta > 0$$

$$\Gamma(z) = \int_0^\infty t^{z-1} c^{-t} dt = (z-1)!$$

$$E[\lambda] = \frac{\alpha}{\alpha+\beta}$$

$$B(p,q) = \frac{q-1}{p+q+1} B(p, q-1)$$

α, β decide the coin fact λ

Categorical Distribution

$$Pr(x = k) = \lambda_k$$

$$Pr(x) = Cat_x[\lambda]$$

Dirichlet Distribution

$$Pr(\lambda_1 \dots \lambda_K) = \frac{\Gamma[\sum_{k=1}^K]\alpha_k}{\prod_{k=1}^K \Gamma[\alpha_k]} \prod_{k=1}^K \lambda_k^{\alpha_k - 1}$$

$$Pr(\lambda_1 \dots \lambda_K) = \mathrm{Dir}_{\lambda_1 \dots \lambda_K}[\alpha_1, \alpha_2 \dots, \alpha_K]$$

Univariate Normal Distribution

$$Pr(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-0.5(x-\mu)^2/\sigma^2]$$
$$Pr(x) = \text{Norm}_x[\mu, \sigma^2]$$

Normal Inverse Gamma Distribution

$$Pr(\mu, \sigma^2) = \frac{\sqrt{\gamma}\beta^\alpha}{\sigma\sqrt{2\pi}\Gamma[\alpha]}(\frac{1}{\sigma^2})^{\alpha+1} \exp[-\frac{2\beta + \gamma(\delta-\mu)^2}{2\sigma^2}]$$
$$Pr(\mu, \sigma^2) = \text{NormInvGam}_{\mu,\sigma^2}[\alpha, \beta, \gamma, \delta]$$

Multivariate Normal Distribution

$$Pr(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\sum|^{1/2}} \exp[-0.5(\mathbf{x}-\mu)^T\sum^-1(\mathbf{x}-\mu)]$$

Normal Inverse Wishart

$$Pr(\mu, \Sigma) = \frac{\gamma^{D/2}|\Psi|^{\alpha/2}|\Sigma|^{-\frac{\alpha+D+2}{2}}}{(2\pi)^{D/2}2^{\frac{\alpha D}{2}}\Gamma_D(\frac{\alpha}{2})} \exp\{-\frac{1}{2}(Tr(\Psi\Sigma^{-1})) + \gamma(\mu-\delta)^T\Sigma^{-1}(\mu-\delta)\}$$

Conjugate Distribution and Conjugate prior

Conjugate Distribution is between prior and posterior

Prior is the conjugate prior of the likelihood function.

2 Fitting model

maximum likelihood

Fitting

$$\begin{aligned}
\hat{\theta} &= argmax_{(\theta)}[Pr(\mathbf{x_{1\ldots I}}|\theta)] \\
&= argmax_{(\theta)}[\prod_{i=1}^{I} Pr(\mathbf{x_i}|\theta)]
\end{aligned}$$

Predictive Density

Evaluate new data point $\mathbf{x}^*$ under probability distribution $Pr(\mathbf{x}^*|\hat{\theta})$ with best parameter.

maximum a posteriori

Fitting

$$\begin{aligned}
\hat{\theta} &= argmax_{(\theta)}[Pr(\theta|\mathbf{x_{1\ldots I}})] \\
&= argmax_{(\theta)}\left[\frac{Pr(\mathbf{x_{1\ldots I}}|\theta)Pr(\theta)}{Pr(\mathbf{x_{1\ldots I}})}\right] \\
&= argmax_{(\theta)}\left[\frac{\prod_{i=1}^{I} Pr(\mathbf{x_i}|\theta)Pr(\theta)}{Pr(\mathbf{x_{1\ldots I}})}\right] \\
\hat{\theta} &= argmax_{(\theta)}\left[Pr(\mathbf{x_i}|\theta)Pr(\theta)\right]
\end{aligned}$$

Predictive

Evaluate new data point $\mathbf{x}^*$ under probability distribution $Pr(\mathbf{x}^*|\hat{\theta})$ with best parameter.

bayesian approach

Fitting

$$Pr(\theta|\mathbf{x_{1\ldots I}}) = \frac{(\prod_{i=1}^{I} Pr(\mathbf{x_i}|\theta))Pr(\theta)}{Pr(\mathbf{x_{1\ldots I}})}$$

The difference between bayesian approach and MAP is that MAP takes the maximum value, while bayesian approach takes the distribution.

Predictive

$$Pr(\mathbf{x}^*|\mathbf{x_{1\ldots I}}) = \int Pr(\mathbf{x}^*|\theta)Pr(\theta|\mathbf{x_{1\ldots I}})d\theta$$

Confusion: the formula should be $\int Pr(\mathbf{x}^*|\theta, \mathbf{x_{1\ldots I}})Pr(\theta|\mathbf{x_{1\ldots I}})d\theta$. Given the θ, it considers $\mathbf{x_{1\ldots I}}$ and x^* are independent.