Machine Vision

October 16, 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

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

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

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

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

Predictive

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

bayesian approach

Fitting

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

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

Predictive

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

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

Multivariate Normal Distribution

If $\mathbf{x_1}, \mathbf{x_2} \dots \mathbf{x_n}$ are independent, the covariance matrix would be

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

Therefore, while $\mathbf{x_1}, \mathbf{x_2} \dots \mathbf{x_n}$ are dependent, the covariance matrix could be decomposed into rotation matrix and diagonal:

$$\Sigma_{full} = \mathbf{R}^T \Sigma'_{diag} \mathbf{R}$$

Marginal Distribution

$$u_i = u_i$$

$$\Sigma_i = \Sigma_{ii}$$

Conditional Distribution

$$u_{i|j} = u_i + \Sigma_{ij}\Sigma_{jj}^{-1}(x_j - u_j)$$

$$\Sigma_{i|j} = \Sigma_{jj} - \Sigma_{ij}^T \Sigma_{ii}^{-1} \Sigma_{ij}$$

Product of two normals

$$\begin{aligned} \mathrm{Norm}_{\mathbf{x}}[\mathbf{a}, \mathbf{A}]\mathrm{Norm}_{\mathbf{x}}[\mathbf{b}, \mathbf{B}] &= k \cdot \mathrm{Norm}_{\mathbf{x}}[(\mathbf{A^{-1}} + \mathbf{B^{-1}})^{-\mathbf{1}}(\mathbf{A^{-1}a} + \mathbf{B^{-1}b}), (\mathbf{A^{-1}} + \mathbf{B^{-1}})^{-\mathbf{1}}] \\ k &= \mathrm{Norm}_{\mathbf{a}}[\mathbf{b}, \mathbf{A} + \mathbf{B}] \end{aligned}$$

change of variables

$$\mathrm{Norm}_{\mathbf{x}}[\mathbf{Ay} + \mathbf{b}, \mathbf{\Sigma}] = k \cdot \mathrm{Norm}_{\mathbf{y}}[\mathbf{A}'\mathbf{x} + \mathbf{b}', \mathbf{\Sigma}']$$

where

$$\begin{aligned} \mathbf{A}' &= \Sigma' A^T \Sigma^{-1} \\ b' &= -\Sigma' A^T \Sigma^{-1} b \\ \Sigma &= (A^T \Sigma^{-1} A)^{-1} \end{aligned}$$

Learning and Inference

The observe measured data, $\mathbf{x}$

Draw inference from it about the state of world, $\mathbf{w}$

If $\mathbf{w}$ is continuous, call this regression.

If $\mathbf{w}$ is discrete, call this classification.

To compute the probability distribution $Pr(\mathbf{w}|\mathbf{x})$, we need: a model(relates visual data $\mathbf{x}$ and $\mathbf{w}$, the relationships depends on parameter θ), a learning algorithm(fits parameter θ from paired training examples $\mathbf{x_i}, \mathbf{w_i}$), an inference algorithm (use model to return $Pr(\mathbf{w}|\mathbf{x})$ given new observed data $\mathbf{x}$)

Types of Model

2.0.1 1. Model contingency of the world on the data $Pr(\mathbf{w}|\mathbf{x})$