1 k-Nearest Neighbour

$$\mathcal{P}\left(k\left|\lambda\right.\right)=\frac{\lambda^{k}e^{-\lambda}}{k!} \tag{1}$$

$$\mathcal{P}\left(k=0\left|D\right.\right)=e^{-2nD}$$

$$\mathcal{P}\left(k=1\left|\Delta D\right.\right)=2n\,e^{-2n\Delta D}\,\Delta D$$

$$\mathcal{P}\left(D\left|n\right.\right)\Delta D=e^{-2nD}\centerdot 2n\,e^{-2n\Delta D}\,\Delta D$$

$$\lim_{\Delta D\to 0}\left(\mathcal{P}\left(D\left|n\right.\right)\Delta D\right)=\lim_{\Delta D\to 0}\left(e^{-2nD}\centerdot 2n\,e^{-2n\Delta D}\,\Delta D\right)$$

$$\mathcal{P}\left(D\left|n\right.\right)=2n\,e^{-2nD} \tag{2}$$

2 MCMC Analysis

$$n\left(\theta\right)=\frac{1}{a}L\left(x\left|\theta\right.\right) \tag{3}$$

$$\mathcal{P}\left(x\right)=\int d\theta\,L\left(x\left|\theta\right.\right)\pi\left(\theta\right) \tag{4}$$

Since we assume a uniform prior,

$$\mathcal{P}\left(x\right)=\frac{a}{\Delta\theta}\int d\theta\,n\left(\theta\right)$$

We have samples from n, with density,

$$\hat{n}\left(\theta\right)=\sum_{\alpha=1}^{S}\delta\left(\theta-\theta_{\alpha}\right) \tag{5}$$

The integral over θ is simply S, that is,

$$\hat{n}\left(\theta\right) = S$$

Hence, the estimate of the Bayesian Evidence is

$$\hat{\mathcal{P}}\left(x\right) = \frac{aS}{\Delta\theta} \tag{6}$$

3 Posterior of Local Density n

We use a Jeffreys prior on the scale parameter n, that is, $\mathcal{P}\left(n\right) \propto \frac{1}{n}$

$$\mathcal{P}\left(n\left|D\right.\right) \propto \mathcal{P}\left(D\left|n\right.\right)\mathcal{P}\left(n\right)$$

$$\mathcal{P}\left(n\left|D\right.\right) \propto 2n, e^{-2nD}\,\frac{1}{n}$$

$$\mathcal{P}\left(n\left|D\right.\right) \propto e^{-2nD} \tag{7}$$

Assumption - $\mathcal{P}\left(D\right)$ is a constant.

4 Posterior Probability of a

We again use a Jeffreys prior on the scale parameter a, that is, $\mathcal{P}\left(a\left|L\right.\right) \propto \frac{1}{a}$

$$\mathcal{P}\left(a\left|D,\, L\right.\right) \propto \mathcal{P}\left(D\left|a,\, L\right.\right)\mathcal{P}\left(a\left|L\right.\right)$$

$$\mathcal{P}\left(a\left|D,\, L\right.\right) \propto \int \mathcal{P}\left(D,\, n\left|a,\, L\right.\right)\mathcal{P}\left(a\left|L\right.\right)\mathcal{P}\left(n\left|a,\, L\right.\right) dn$$

We assume $\mathcal{P}\left(n\left|a,\, L\right.\right) = \delta\left(n - \frac{L}{a}\right)$. Therefore,

$$\mathcal{P}\left(a\left|D,\, L\right.\right) \propto \int \mathcal{P}\left(D,\, n\left|a,\, L\right.\right)\,\frac{1}{a}\,\delta\left(n - \frac{L}{a}\right) dn$$

But $\mathcal{P}\left(D,\, n\left|a,\, L\right.\right) \propto \mathcal{P}\left(n\left|D\right.\right)$

$$\mathcal{P}\left(a\,|D,\,L\right)\propto\int e^{-2nD}\,\frac{1}{a}\,\delta\left(n-\frac{L}{a}\right)dn$$

$$\mathcal{P}\left(a\,|D,\,L\right)\propto\frac{1}{a^{2}}\exp\left(-\frac{2DL}{a}\right)$$

5 Considering All Samples

$$\mathcal{P}\left(a\,|D,\,L\right)=k\,\left(\prod_{\alpha=1}^{S}\frac{1}{a}\exp\left(-\frac{2D_{\alpha}L_{\alpha}}{a}\right)\right)\,\frac{1}{a}$$

$$\log\mathcal{P}\left(a\,|D,\,L\right)=\log k+\log\left(\prod_{\alpha=1}^{S}\frac{1}{a}\right)+\log\left(\prod_{\alpha=1}^{S}\exp\left(-\frac{2D_{\alpha}L_{\alpha}}{a}\right)\right)+\log\left(\frac{1}{a}\right)$$

$$\log\mathcal{P}\left(a\,|D,\,L\right)=\log k-S\log a-\frac{2}{a}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha}-\log a$$

$$\log\mathcal{P}\left(a\,|D,\,L\right)=\mathrm{const}-\left(S+1\right)\,\log a-\frac{2}{a}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha}\tag{8}$$

5.1 Most Probable Value of a

$$\frac{d}{da}\left[\mathrm{const}-\left(S+1\right)\,\log a-\frac{2}{a}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha}\right]=-\frac{\left(S+1\right)}{a}+\frac{2}{a^{2}}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha}$$

$$\frac{2}{a^{2}}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha}=\frac{S+1}{a}$$

$$a=\frac{2}{S+1}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha}\tag{9}$$

In general,

$$a=\frac{1}{Sk+1}\sum_{\alpha}V_{d}\left(D_{\alpha}\right)L_{\alpha}$$

5.2 Estimate of the Bayesian Evidence

$$\hat{\mathcal{P}}\left(x\right)=\frac{aS}{\Delta\theta}$$

$$\hat{\mathcal{P}}\left(x\right)=\frac{2S}{\Delta\theta\left(S+1\right)}\sum_{\alpha=1}^{S}D_{\alpha}L_{\alpha} \tag{10}$$

6 Bayesian Evidence

We have $\mathbf{x}=\begin{pmatrix}x_1^{(i)}\\x_2^{(i)}\end{pmatrix}$ and we want to find $\boldsymbol{\mu}=\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}$. We will assume uniform priors on μ_1 and μ_2 such that $\mathcal{P}\left(\mu_1\right)=\dfrac{1}{\Delta\mu_1}$ and $\mathcal{P}\left(\mu_2\right)=\dfrac{1}{\Delta\mu_2}$. Using Bayes' Theorem,

$$\mathcal{P}\left(\boldsymbol{\mu}\left|\mathbf{x}\right.\right)=\frac{\mathcal{P}\left(\mathbf{x}\left|\boldsymbol{\mu}\right.\right)\mathcal{P}\left(\boldsymbol{\mu}\right)}{\mathcal{P}\left(\mathbf{x}\right)} \tag{11}$$

where

$$\mathcal{P}\left(\mathbf{x}\left|\boldsymbol{\mu}\right.\right)=\frac{1}{\left|2\pi\Sigma\right|^{\frac{N}{2}}}\exp\left[-\frac{1}{2}\sum_i\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)\right]$$

The Bayesian Evidence is given by

$$\mathcal{P}\left(\mathbf{x}\right)=\int\mathcal{P}\left(\mathbf{x}\left|\boldsymbol{\mu}\right.\right)\mathcal{P}\left(\boldsymbol{\mu}\right)d\boldsymbol{\mu} \tag{12}$$

Assuming the priors to be independent,

$$\mathcal{P}\left(\mathbf{x}\right)=\int\mathcal{P}\left(\mathbf{x}\left|\boldsymbol{\mu}\right.\right)\mathcal{P}\left(\mu_1\right)\mathcal{P}\left(\mu_2\right)d\boldsymbol{\mu}$$

$$\mathcal{P}\left(\mathbf{x}\right)=\frac{1}{\Delta\mu_1\Delta\mu_2}\int\mathcal{P}\left(\mathbf{x}\left|\boldsymbol{\mu}\right.\right)d\boldsymbol{\mu}$$

$$\mathcal{P}\left(\mathbf{x}\right)=\frac{1}{\Delta\mu_1\Delta\mu_2\left|2\pi\Sigma\right|^{\frac{N}{2}}}\int\exp\left[-\frac{1}{2}\sum_i\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)\right]d\boldsymbol{\mu}$$

$$\sum_i\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)=N\boldsymbol{\mu}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\mu}-2\boldsymbol{\mu}^{\mathrm{T}}\sum_i\Sigma^{-1}\mathbf{x}_i+\sum_i\mathbf{x}_i^{\mathrm{T}}\Sigma^{-1}\mathbf{x}_i$$

$\mathbf{A}=N\Sigma^{-1}$

$\mathbf{B}=-2\sum_i\Sigma^{-1}\mathbf{x}_i$

$\mathbf{C} = \sum_i \mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_i$

$$\sum_i \left(\mathbf{x}^{(i)} - \boldsymbol{\mu}\right)^{\mathrm{T}} \Sigma^{-1} \left(\mathbf{x}^{(i)} - \boldsymbol{\mu}\right) = \boldsymbol{\mu}^{\mathrm{T}} \mathbf{A} \boldsymbol{\mu} + \boldsymbol{\mu}^{\mathrm{T}} \mathbf{B} + \mathbf{C}$$

$\boldsymbol{\mu}^{\mathrm{T}} \mathbf{A} \boldsymbol{\mu} + \boldsymbol{\mu}^{\mathrm{T}} \mathbf{B} + \mathbf{C} = (\boldsymbol{\mu} - h)^{\mathrm{T}} \mathbf{A} (\boldsymbol{\mu} - h) + k$
$h = -\dfrac{1}{2} \mathbf{A}^{-1} \mathbf{B} = \dfrac{1}{N} \left(\sum_i \mathbf{x}_i\right)$
$k = \mathbf{C} - \dfrac{1}{4} \mathbf{B}^{\mathrm{T}} \mathbf{A}^{-1} \mathbf{B} = \sum_i \mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_i - \dfrac{1}{N} \left(\sum_i \Sigma^{-1} \mathbf{x}_i\right)^{\mathrm{T}} \Sigma \left(\sum_i \Sigma^{-1} \mathbf{x}_i\right)$

$$\mathcal{P}(\mathbf{x}) = \frac{1}{\Delta\mu_1 \Delta\mu_2 \left|2\pi\Sigma\right|^{\frac{N}{2}}} \int \exp\left[-\frac{1}{2}\left((\boldsymbol{\mu} - h)^{\mathrm{T}} \mathbf{A} (\boldsymbol{\mu} - h) + k\right)\right] d\boldsymbol{\mu}$$

$$\mathcal{P}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}k\right) \left|\frac{2\pi}{N}\Sigma\right|^{\frac{1}{2}}}{\Delta\mu_1 \Delta\mu_2 \left|2\pi\Sigma\right|^{\frac{N}{2}}}$$

$$\mathcal{P}(\mathbf{x}) = \frac{\left|\frac{2\pi}{N}\Sigma\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\sum_i \mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_i - \frac{1}{N}\left(\sum_i \Sigma^{-1} \mathbf{x}_i\right)^{\mathrm{T}} \Sigma \left(\sum_i \Sigma^{-1} \mathbf{x}_i\right)\right)\right]}{\Delta\mu_1 \Delta\mu_2 \left|2\pi\Sigma\right|^{\frac{N}{2}}} \tag{13}$$

We will also assume that

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

$\left(\Sigma^{-1}\right)^{\mathrm{T}} = \Sigma^{-1}$

$$\mathcal{P}(\mathbf{x}) = \frac{\left|\frac{2\pi}{N}\Sigma\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\sum_i \mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_i - \frac{1}{N}\left(\sum_i \mathbf{x}_i\right)^{\mathrm{T}} \left(\Sigma^{-1}\right) \left(\sum_i \mathbf{x}_i\right)\right)\right]}{\Delta\mu_1 \Delta\mu_2 \left|2\pi\Sigma\right|^{\frac{N}{2}}} \tag{14}$$

$$\mathcal{P}(\mathbf{x}) = \frac{\left|\frac{2\pi}{N}\Sigma\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_i \left(\mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_i - \frac{1}{N} \mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_i\right)\right]}{\Delta\mu_1 \Delta\mu_2 \left|2\pi\Sigma\right|^{\frac{N}{2}}} \tag{15}$$

7 Gibbs Sampling

$$\mathcal{P}\left(\boldsymbol{\mu}\,|\mathbf{x}\right) \propto \exp\left[-\frac{1}{2}\sum_{i}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}\right)\right]$$

$$\mathcal{P}\left(\boldsymbol{\mu}\,|\mathbf{x}\right) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i}\left(\mathbf{x}_1^{(i)}-\mu_1\right)^2\right]\exp\left[-\frac{1}{2\sigma^2}\sum_{i}\left(\mathbf{x}_2^{(i)}-\mu_2\right)^2\right]$$

$$\mathcal{P}\left(\mu_1\,|\mathbf{x},\,\mu_2\right) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left[-\frac{1}{2\left(\sigma^2/N\right)}\left(\mu_1-\bar{\mathbf{x}}_1\right)^2\right]$$

$$\mathcal{P}\left(\mu_2\,|\mathbf{x},\,\mu_1\right) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left[-\frac{1}{2\left(\sigma^2/N\right)}\left(\mu_2-\bar{\mathbf{x}}_2\right)^2\right]$$