Let $x^{(1)}, \ldots, x^{(n)} \overset{IID}{\sim} P$

$x^{(1)}, \ldots, x^{(n)} \overset{IID}{\sim} N(\mu, \sigma^2)$

$x^{(1)}, \ldots, x^{(n)} \overset{IID}{\sim} Ber(\theta)$

$X \sim N(\mu, \sigma^2) \quad f(X=x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$X \sim Ber(\theta) \quad P(X=1) = \theta, \quad P(X=0) = 1-\theta$

$\mathcal{D} = \{x^{(1)}, \ldots, x^{(n)}\}$ find parameters

$\hat{\theta} = \frac{n_1}{n}$ for $Ber(\theta)$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2$

$E(\theta) = \frac{E(n_1)}{n} = \frac{1}{n}\sum_{i=1}^{n} E(\mathbb{1}(x^{(i)} = 1)) = \frac{n\theta}{n} = \theta$

$E(\bar{x}) = \mu, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2$

$= \frac{1}{n}\left(\sum_{i=1}^{n} z_i^2 - \bar{z}^2 n\right)$

$x_i = z_i + \mu$

$\bar{x} = \bar{z} + \mu$

$E(S^2) = \frac{\sigma^2 n}{n} = \sigma^2$

Maximum Likelihood

$$\mathcal{L}_n(\theta) = \sum_{i=1}^{n} \log p(X^{(i)} = x^{(i)} | \theta)$$

$$\theta_n = \max_{\theta} \mathcal{L}_n(\theta)$$

$$g(t) = t - \log t$$

$$\min_{t} g(t) = ?$$

$$g'(t) = 1 - \frac{1}{t}$$

$$g''(t) = \frac{1}{t^2}.$$

$g(t)$ is convex over $t > 0$

$g'(1) = 0$ and hence it is the minimum

$g(1) \le t - \log t \qquad t > 0$

$1 \le t - \log t$

$l \quad t \le t - 1, \qquad t > 0,$

Let $p_i > 0, \quad \sum_{i=1}^{d} p_i = 1$

$q_i > 0 \quad \sum_{i=1}^{d} q_i = 1$

$P(X = a_i) = p_i$

or

$P(X = a_i) = q_i$

$$\log \frac{q_i}{p_i} \le \frac{q_i}{p_i} - 1$$

$$\sum_{i=1}^{d} p_i \log \frac{q_i}{p_i} \le \sum_{i=1}^{d} q_i - \sum_{i=1}^{d} p_i = 0$$

$$KL(P,Q) = \sum_{i=1} P_i \log \frac{P_i}{q_i}$$

$$KL(P,Q) \geq 0$$

Equality holds iff $P = Q$

P is X with p.m.f $\{P_1, \ldots, P_d\}$

Q is X with p.m.f $\{q_1, \ldots, q_d\}$

$KL(P,Q) = 0$ iff $P_i = q_i$

$i = \{1, \ \ , d\}$

$X \sim Ber(\theta)$ $\mathcal{D} = \{x^{(1)}, \ldots, x^{(n)}\}$

$$p(X = x|\theta) = \theta^x (1-\theta)^{(1-x)}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \left(x^{(i)} \log \theta + (1 - x^{(i)}) \log (1-\theta) \right)$$

$$\mathcal{L}(\theta) = n_1 \log \theta + n_0 \log (1-\theta)$$

$$n_1 = \sum x^{(i)} \quad , \quad n_0 = \sum_{i=1}^{n} \left(1 - x^{(i)}\right)$$

$$\hat{\theta} = \frac{n_1}{n} \qquad 1 - \hat{\theta} = \frac{n_0}{n}.$$

$$n = n_1 + n_0$$

$$\mathcal{L}(\theta) = n\left(\hat{\theta} \log \theta + (1-\hat{\theta}) \log (1-\theta)\right)$$

$$\mathcal{L}(\theta) = n\left(\hat{\theta} \log \frac{\theta}{\hat{\theta}} + (1-\hat{\theta}) \log \frac{(1-\theta)}{(1-\hat{\theta})}\right)$$

$$+ n\left(\hat{\theta} \log \hat{\theta} + (1-\hat{\theta}) \log (1-\hat{\theta})\right)$$

$$S(P) = -\sum_{i=1}^{d} q_i \log q_i \quad \Big| \text{ Entropy}$$

$$\mathcal{L}(\theta) = -n \, KL(P, Q(\theta))$$
$$- n \, S(P)$$

$d = 2, \quad a_1 = 1, \quad a_2 = 0$

$\underline{P}$ $\quad P(X=1) = \hat{\theta} \qquad P(X=0) = (1-\hat{\theta})$

$\underline{Q}$ $\quad P(X=1) = \theta \qquad P(X=0) = 1-\theta$

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta)$$

$$= \underset{\theta}{\operatorname{argmin}}\, KL(P, Q(\theta))$$

$KL(P, Q(\theta)) = 0 \quad$ iff

$$\theta_n = \hat{\theta}$$

$$X \sim N(\mu, C), \quad C \succ 0$$

$$P(X = x \mid \theta) = \frac{1}{(\sqrt{2\pi})^d |C|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)} \qquad |C| = \det(C)$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log P(X = x \mid \theta)$$

$$= \sum_{i=1}^{n} \left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2} \log |C| - \frac{1}{2}(x^{(i)} - \mu)^T C^{-1}(x^{(i)} - \mu) \right)$$

For a fixed C

$$f(\mu) = \frac{1}{2} \sum_{i=1}^{n} (x^{(i)} - \mu)^T C^{-1}(x^{(i)} - \mu)$$

$$\nabla f(\mu) = C^{-1} \sum_{i=1}^{n} (\mu - x^{(i)})$$

$$H(\mu) = n C^{-1}$$

$$f(\mu) \geq f(\mu^*) \qquad C \succ 0$$

$$\text{If } \nabla f(\mu^*) = 0$$

$$\mu^* = \frac{1}{n} \sum x^{(i)} = \bar{x}$$

$$\mathcal{L}(\mu, C) \leq \mathcal{L}(\bar{x}, C)$$

$$S = \frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^T$$

$$\mathcal{L}(\mu, C)$$

$$= \sum_{i=1}^{n} \left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2} \log |C| - \frac{1}{2} (x^{(i)} - \mu)^T C^{-1} (x^{(i)} - \mu) \right)$$

-

$$= n\left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2}\log|C|\right)$$

$$-\frac{1}{2}\mathrm{Tr}\left(C^{-1}\sum_{i=1}^{n}(x^{(i)}-\mu)(x^{(i)}-\mu)^T\right)$$

$$\mathcal{L}(\bar{x}, S) = n\left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2}\log|S|\right)$$

$$-\frac{1}{2}\mathrm{Tr}\left(S^{-1}\sum_{i=1}^{n}(x^{(i)}-\bar{x})(x^{(i)}-\bar{x})^T\right)$$

$$= n\left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2}\log|S|\right)$$

$$-\frac{n}{2}\mathrm{Tr}\left(S^{-1}S\right)$$

$$= n\left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2}\log|S|\right) - \frac{n}{2}d$$

$\mathcal{L}(\bar{x}, C) =$

$$n\left(\log \frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2}\log|C|\right) -$$

$$\frac{1}{2}\mathrm{Tr}\left(C^{-1}\sum_{i=1}^{n} (x^{(i)}-\bar{x})(x^{(i)}-\bar{x})^T\right)$$

$$= n\left(\log\frac{1}{(\sqrt{2\pi})^d} - \frac{1}{2}\log|C|\right)$$

$$- \frac{n}{2}\mathrm{Tr}(C^{-1}S)$$

$$\mathcal{L}(\bar{x}, S) - \mathcal{L}(\bar{x}, C)$$

$$= -\frac{n}{2}\log|S| + \frac{n}{2}\log|C| - \frac{nd}{2}$$

$$+ \frac{n}{2}\mathrm{Tr}(C^{-1}S)$$

$$= \frac{n}{2}\left(-\log\frac{|S|}{|C|} - d + \mathrm{Tr}(C^{-1}S)\right)$$

$$A = C^{-1}S$$

$$\det(A) = \frac{\det(S)}{\det(C)}.$$

$$\sum_{i=1}^{d} \lambda_i(A) = \det(A)$$

$$\mathrm{Trace}(A) = \sum_{i=1}^{d} \lambda_i(A)$$

$$\mathcal{L}(\bar{X}, S) - \mathcal{L}(\bar{X}, C)$$

$$= \frac{n}{2}\left(-\log|A| - d + \mathrm{Tr}(A)\right)$$

$$= \frac{n}{2}\sum_{i=1}^{d}\left(-\log\lambda_i(A) - 1 + \lambda_i(A)\right)$$

$$\geq 0$$

Equality holds iff $\lambda_i(A) = 1$

$$= A = I$$

$$\therefore \mathcal{L}(\mu, C) \le \mathcal{L}(\bar{x}, C) \le \mathcal{L}(\bar{x}, S)$$