Assignment 5

1. Given

$$f(x) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)},$$

where $x, \mu \in \mathbb{R}^k$, Σ is a k-by-k positive definite matrix and $|\Sigma|$ is its determinant. Show that $\int_{\mathbb{R}^k} f(x)\,dx = 1$.

Σ is a positive definite matrix, then $\Sigma = Q\Lambda Q^T$, where $Q^TQ = I$ and $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_k \end{bmatrix}$ (λ_i are eigenvalues)

Define $\Sigma^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}Q^T$, where $\Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \sqrt{\lambda_2} & \\ 0 & & \ddots \ \sqrt{\lambda_k} \end{bmatrix}$ and

$\Sigma^{-\frac{1}{2}} = Q\Lambda^{-\frac{1}{2}}Q^T$, where $\Lambda^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & \\ & \frac{1}{\sqrt{\lambda_2}} & \\ & & \ddots \ \frac{1}{\sqrt{\lambda_k}} \end{bmatrix}$

Then $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}Q^TQ\Lambda^{\frac{1}{2}}Q^T = Q\Lambda Q^T = \Sigma$ and

$\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}} = Q\Lambda^{-\frac{1}{2}}Q^TQ\Lambda^{-\frac{1}{2}}Q^T = Q\Lambda^{-1}Q^T = \Sigma^{-1}$

Let $y = \Sigma^{-\frac{1}{2}}(x-\mu)$.

Then $x = \Sigma^{\frac{1}{2}}y + \mu$ and $dx = |\Sigma^{\frac{1}{2}}|\,dy = |\Sigma|^{\frac{1}{2}}\,dy$.

$$\Rightarrow f(x)\,dx = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}(\Sigma^{\frac{1}{2}}y)^T\Sigma^{-1}(\Sigma^{\frac{1}{2}}y)} |\Sigma|^{\frac{1}{2}}\,dy$$

$$= \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}y^Ty} |\Sigma|^{\frac{1}{2}}\,dy = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}y^Ty}\,dy.$$

$$\Rightarrow \int_{\mathbb{R}^k} f(x)\,dx = \int_{\mathbb{R}^k} \frac{1}{(2\pi)^{k/2}} e^{\frac{-1}{2}y^Ty}\,dy \,,$$

where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$, and $y^Ty = y_1^2 + y_2^2 + \cdots + y_k^2$

$$= \int_{\mathbb{R}^k} \frac{1}{(2\pi)^{k/2}} e^{\frac{-1}{2}\sum_{i=1}^{k} y_i^2}\,dy$$

$$= \int_{\mathbb{R}^k} \frac{1}{(2\pi)^{k/2}} \prod_{i=1}^{k} e^{\frac{-1}{2}y_i^2}\,dy \,.$$

$$= \frac{1}{(2\pi)^{k/2}} \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{\frac{-1}{2}y_i^2}\,dy_i$$

$$= \frac{1}{(2\pi)^{k/2}} \cdot (\sqrt{2\pi})^k = 1 \,.$$

2. Let A, B be n-by-n matrices and x be a n-by-1 vector.
   (a) Show that $\frac{\partial}{\partial A}\mathrm{trace}(AB) = B^T$.
   (b) Show that $x^TAx = \mathrm{trace}(xx^TA)$.
   (b) Derive the maximum likelihood estimators for a multivariate Gaussian.

(a) $\mathrm{trace}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i,j} A_{ij}B_{ji}$

$$\frac{\partial}{\partial A_{pq}} \mathrm{trace}(AB) = \frac{\partial}{\partial A_{pq}} \sum_{i,j} A_{ij}B_{ji} = B_{qp} = (B^T)_{pq}$$

Hence $\frac{\partial}{\partial A}\mathrm{trace}(AB) = B^T$.

(b) $x^TAx = \sum_{i,j} x_i A_{ij} x_j$

$\downarrow$ since x^TAx is a scalar

$= \mathrm{trace}(\underline{x^TA}\,x)$

$\downarrow$ trace AB = trace BA, whenever AB and BA are squares

$= \mathrm{trace}(x\,\underline{x^TA})$

(c) Given data $\{x^{(i)}\}_{i=1}^{m}$ with $x^{(i)} \sim N(\mu, \Sigma)$

$$L(\mu,\Sigma) = \prod_{i=1}^{m} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left[-\frac{1}{2}(x^{(i)}-\mu)^T \Sigma^{-1} (x^{(i)}-\mu)\right]$$

$$\Rightarrow \ell(\mu,\Sigma) = \frac{-m}{2} \log|2\pi\Sigma| - \frac{1}{2}\sum_{i=1}^{m} (x^{(i)}-\mu)^T \Sigma^{-1} (x^{(i)}-\mu)$$

For $f = (x-\mu)^T A (x-\mu)$, $\frac{\partial f}{\partial \mu} = -2A(x-\mu)$

$$\Rightarrow \frac{\partial \ell}{\partial \mu} = \Sigma^{-1} \sum_{i=1}^{m} (x^{(i)}-\mu)$$

When $\frac{\partial \ell}{\partial \mu} = 0$, $\sum_{i=1}^{m}(x^{(i)}-\mu) = 0 \Rightarrow \underline{\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}}$.

$\frac{\partial}{\partial \Sigma^{-1}} \log|\Sigma| = -\Sigma$ and $\frac{\partial}{\partial \Sigma^{-1}} \operatorname{trace}(\Sigma^{-1} S) = S^T = S$.

Take $S = \sum_{i=1}^{m} (x^{(i)}-\mu)(x^{(i)}-\mu)^T$

$$\Rightarrow \frac{\partial \ell}{\partial \Sigma^{-1}} = \frac{m}{2}\Sigma - \frac{1}{2} S$$

When $\frac{\partial \ell}{\partial \Sigma^{-1}} = 0$, $\Sigma = \frac{1}{m} S \Rightarrow \underline{\hat{\Sigma} = \frac{1}{m}\sum_{i=1}^{m} (x^{(i)}-\hat{\mu})(x^{(i)}-\hat{\mu})^T}$

3. Unanswered Questions

   There are unanswered questions from the lecture, and there are likely more questions we haven't covered.

- Take a moment to think about these questions.
- Write down the ones you find important, confusing, or interesting.
- You do **not** need to answer them—just state them clearly.

What factors affect the performance of softmax regression ?

eg. label imbalance ? sample size ?