Tutorial 4

PS 4-12 Lagrange Multipliers & equality constraints

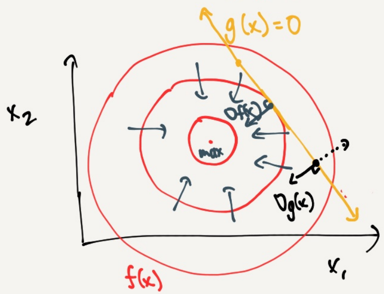
$$\{\pi_j\} = \operatorname*{argmax}_{\pi_j} \sum_j z_j \log \pi_j \quad , \quad \underbrace{\text{s.t. } \sum_{j=1}^{K} \pi_j = 1 \quad , \quad \pi_j \geq 0}$$

Generic opt. problem w/ equality constraints.

$$x^* = \operatorname*{argmax}_x f(x) \leftarrow \text{objective}$$

$$\text{s.t. } g(x) = 0$$

↖ constraint



Note the following:

1) $\nabla g(x)$ is orthogonal to the constraint surface b/c $g(x)$ is constant on the surface $g(x)=0$.

2) at x^*, $\nabla f(x)$ is orthogonal to the constraint surface. (otherwise, $f(x^*)$ could be increased.)

Thus $\nabla f(x)$ & $\nabla g(x)$ should be parallel or anti-parallel, i.e.

$$\nabla f(x) + \lambda \nabla g(x) = 0 \quad , \quad \lambda \neq 0$$

Define the Lagrangian function:

$$L(x,\lambda) = f(x) + \lambda g(x)$$

↖ Lagrange multiplier

$$\frac{\partial L}{\partial x} = \nabla f(x) + \lambda \nabla g(x) = 0 \Leftarrow \text{optimality condition}$$

$$\frac{\partial L}{\partial \lambda} = g(x) = 0 \Leftarrow \text{constraint}$$

Hence, solving for $\frac{\partial L}{\partial x} = 0$ & $\frac{\partial L}{\partial \lambda} = 0$ will give x^*.

a) $$\operatorname*{argmax}_{\pi_j} \underbrace{\sum_j z_j \log \pi_j}_{f}$$

$$\text{s.t. } \underbrace{\sum_j \pi_j = 1} \rightarrow \underbrace{\sum_j \pi_j - 1}_{g} = 0$$

Lagrangian:

$$L(\pi,\lambda) = \sum_j z_j \log \pi_j + \lambda\left(\sum_j \pi_j - 1\right)$$

$$\frac{\partial L}{\partial \lambda} = \sum_j \pi_j - 1 = 0 \Rightarrow \sum_j \pi_j = 1$$

$$\frac{\partial L}{\partial \pi_j} = \frac{z_j}{\pi_j} + \lambda = 0 \Rightarrow z_j + \lambda \pi_j = 0$$

sum over j: $$\sum_j (z_j + \lambda \pi_j) = 0$$

$$\sum_j z_j + \lambda \underbrace{\sum_j \pi_j}_{1} = 0 \Rightarrow \lambda = -\sum_j z_j$$

$$z_j + \lambda \pi_j = 0$$

$$\pi_j = \frac{-z_j}{\lambda}$$

$$\boxed{\pi_j^* = \frac{z_j}{\sum_k z_k}}$$

CS5487 Lecture Notes (2022B)
Prof. Antoni B. Chan
Dept of Computer Science
City University of Hong Kong

4-6 Mixture of exponentials

$$p(x) = \sum_{j=1}^{K} \underbrace{\pi_j}_{\text{prior/component prob.}} p(x|j) \quad , \quad \underbrace{p(x|j) = \lambda_j e^{-\lambda_j x}}_{\text{exponential component}}$$

Given dataset $D = \{x_1, \dots, x_n\}$.

joint LL: $\log p(X, Z) = \sum_i \log p(x_i, z_i) = \sum_i \log p(x_i|z_i)p(z_i)$

indicator variable trick: $= \sum_i \sum_j z_{ij} \log \pi_j + z_{ij} \log p(x_i|z_i=j)$

E-step : Q-function:

$$Q(\theta;\hat{\theta}) = E_{Z|X,\hat{\theta}}[\log p(X,Z)]$$

$$= \sum_i \sum_j \underbrace{E_{Z|X}[z_{ij}]}_{\hat{z}_{ij}} \log \pi_j + \underbrace{E_{Z|X}(z_{ij})}_{\hat{z}_{ij}} \log p(x_i|z_i=j)$$

$$\hat{z}_{ij} = E_{Z|X}[z_{ij}] = \frac{\pi_j \, p(x_i|z_i=j)}{\sum_k \pi_k \, p(x_i|z_i=k)} \quad \text{(from lecture)}$$

P1-5 : expectation of indicator:

$x \in \{0,1\}$, $p(x)$

$$E[x] = \sum_{x\in\{0,1\}} p(x=1)\cdot x = p(x=1)\cdot 1 + p(x=0)\cdot 0$$
$$= p(x=1)$$

M-step

π_j : $\{\hat{\pi}_j\} = \underset{\pi_j}{\text{argmax}} \underbrace{\sum_i \sum_j}\hat{z}_{ij} \log \pi_j$, s.t. $\sum_j \pi_j = 1$, $\pi_j \geq 0$

$$= \underset{\pi_j}{\text{argmax}} \sum_j \underbrace{\left(\sum_i \hat{z}_{ij}\right)}_{\hat{N}_j} \log \pi_j \quad , \dots$$

From (4-12),

$$\hat{\pi}_j = \frac{\hat{N}_j}{\sum_k \hat{N}_k}$$

λ_j: $\hat{\lambda}_j = \underset{\lambda_j}{\text{argmax}} \sum_i \sum_k \hat{z}_{ik} \log p(x_i|z_i=k)$ ignore all $k \neq j$ since they don't have λ_j.

$$= \underset{\lambda_j}{\text{argmax}} \sum_i \hat{z}_{ij} \log p(x_i|z_i=j)$$

$$= \underset{\lambda_j}{\text{argmax}} \sum_i \hat{z}_{ij} \log(\lambda_j e^{-\lambda_j x_i})$$

$$= \underset{\lambda_j}{\text{argmax}} \sum_i \hat{z}_{ij} (\log \lambda_j - \lambda_j x_i)$$

$$= \underset{\lambda_j}{\text{argmax}} \underbrace{\left(\sum_i \hat{z}_{ij}\right)}_{\hat{N}_j} \log \lambda_j - \lambda_j \sum_i \hat{z}_{ij} x_i$$

$$\frac{\partial}{\partial \lambda_j} = \frac{\hat{N}_j}{\lambda_j} - \sum_i \hat{z}_{ij} x_i = 0$$

$$\Rightarrow \boxed{\frac{1}{\lambda_j} = \frac{1}{\hat{N}_j} \sum_i \hat{z}_{ij} x_i}$$