Homework #4

Simas Glinskis

December 1, 2016

Collaborated with Natasha Antropova

Problem 1

We need to show that the two training objectives are equivalent. Let's call,

$$A = \sum_{i=1}^{N} (y_i - (w \cdot \phi(x_i) + w_0))^2 + \lambda ||w||^2$$

$$B = \sum_{i=1}^{N} (y_i - (w \cdot \phi(x_i) + w_0))^2 + \alpha(||w||^2 - \tau)$$

Where we have used a Lagrange multiplier, α, for the constraint.

For the optimal solutions, we know:

$$\nabla_w A(w_\lambda^*) = 0$$

and using the KKT conditions,

$$\nabla_w B(w_\alpha^*) = \nabla_w A(w_\alpha^*) = 0 \text{ and } \alpha(||w_\alpha^*||^2 - \tau) = 0$$

where w_λ^* are the optimal weights for a given λ and w_α^* are the optimal weights for a given α.

Say we set $\tau = ||w_\lambda^*||^2$, then we have $\lambda = \alpha$ and $w_\lambda^* = w_\alpha^*$, satisfying the KKT condition.

Therefore, the optimal weights will be the same for a $\tau = ||w_\lambda^*||^2$.

Problem 2

See attached notebook.

Problem 3

The posterior probability $\gamma_{i,c} = p(z_i = c|x_i; \theta, \pi)$ may be written as,

$$\gamma_{i,c} = \frac{\pi_c p(x_i|\theta_c)}{\sum_{l=1}^{k} \pi_l p(x_i|\theta_l)}$$

For a Bernoulli mixture,

$$p(x_i|\theta_c) = \prod_{j=1}^{d} \theta_{c,j}^{x_{i,j}} (1 - \theta_{c,j})^{(1-x_{i,j})}$$

Therefore,

$$\gamma_{i,c} = \frac{\pi_c \prod_{j=1}^{d} \theta_{c,j}^{x_{i,j}} (1 - \theta_{c,j})^{(1-x_{i,j})}}{\sum_{l=1}^{k} \pi_l \prod_{j=1}^{d} \theta_{l,j}^{x_{i,j}} (1 - \theta_{l,j})^{(1-x_{i,j})}}$$

Problem 4

The M-step updates are,

$$\pi^{new}, \theta^{new} = argmax_{\pi,\theta} \big[\sum_{i=1}^{N} \sum_{c=1}^{k} \gamma_{i,c} (\log \pi_c + \log p(x_i; \theta_c)) \big]$$

Let us set,

$$A = \sum_{i=1}^{N} \sum_{c=1}^{k} \gamma_{i,c} (\log \pi_c + \log p(x_i; \theta_c)) = \sum_{i=1}^{N} \sum_{c=1}^{k} \gamma_{i,c} (\log \pi_c + \sum_{j=1}^{d} [x_{i,j} \log \theta_{c,j} + (1 - x_{i,j}) \log(1 - \theta_{c,j})])$$

Solving for π^{new} using Lagrange multipliers and differentiating,

$$A = \sum_{i=1}^{N} \sum_{c=1}^{k} \gamma_{i,c} (\log \pi_c) + \alpha (\sum_{c=1}^{k} \pi_c - 1) + ...$$

$$\frac{\partial A}{\partial \pi_c} = 0 \rightarrow \sum_{i=1}^{N} \gamma_{i,c} = -\alpha \pi_c$$

$$\sum_{c=1}^{k} \sum_{i=1}^{N} \gamma_{i,c} = -\sum_{c=1}^{k} \alpha \pi_c \rightarrow \sum_{i=1}^{N} 1 = -\alpha \rightarrow \alpha = -N$$

$$\pi_k^{new} = \frac{1}{N} \sum_{i=1}^{N} \gamma_{i,k}$$

for a given subclass k.

Solving for θ^{new},

$$\frac{\partial A}{\partial \theta_{k,r}} = 0 = \sum_{i=1}^{N} \gamma_{i,k} \Big(\frac{x_{i,r}}{\theta_{k,r}} - \frac{1 - x_{i,r}}{1 - \theta_{k,r}} \Big) = \sum_{i=1}^{N} \gamma_{i,k} (x_{i,r} - \theta_{k,r}) = 0$$

$$\theta_{k,r}^{new} = \frac{\sum_{i=1}^{N} \gamma_{i,k} x_{i,r}}{\sum_{i=1}^{N} \gamma_{i,k}}$$

for a given subclass k and parameter r.

Problem 5

See attached notebook.