Expected value of gamma distribution

$$\mathbb{E}[y] = \int_0^\infty y \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}\, dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \underbrace{y^\alpha e^{-\beta y}}_{\text{gamma}(\alpha+1,\beta)} dy$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \underbrace{\int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} y^\alpha e^{-\beta y}\, dy}$$

this is now equal to the gamma distribution

$P(y \mid \alpha+1, \beta)$ → has to equal 1

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} = \cancel{\beta} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta}$$

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u}\, du \qquad \Gamma(\alpha+1) = \int_0^\infty u^\alpha e^{-u}\, du$$

gamma function is the continuous analog of the parameter function $(n-1)!$

thus we can rewrite

$$\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\alpha!}{(\alpha-1)!} = \alpha$$

$$= \frac{\alpha}{\beta}$$

HW5

take l_2 norm

$$\left\| x_i - \sum_{j=1}^{k} z_{ij} v_j \right\|_2^2 = \left(x_i - \sum_{j=1}^{k} z_{ij} v_j \right)^T \left(x_i - \sum_{j=1}^{K}\right.$$

$$= x_i^T x_i - \sum_{j=1}^{k} z_{ij} v_j^T x_i - x_i^T$$

$$= x_i^T x_i - 2\sum_{j=1}^{k} z_{ij} v_j^T x_i + \sum_{j=1}^{k} v$$

$$\cancel{= x_i^T x_i \sum}$$

$$= x_i^T x_i - \sum_{j=1}^{} v_j^T x_i x_i^T v_j$$

$$\cancel{J_k = \frac{1}{n} \left(x_i^T x_i - \sum_{j=1} x_i^T \ldots \sum_j \right)}$$

$$J_k = \frac{1}{n} \sum_{i=1}^{n} \left(x_i^T x_i - \sum_{j=1}^{k} v_j^T x_i x_i^T v_j \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} \lambda_j$$

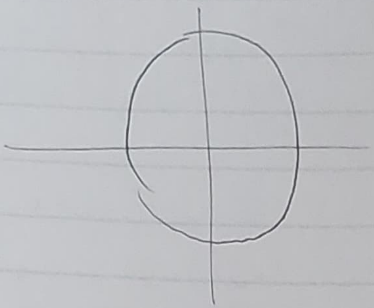
unsure - checked answer key

$$J_d = 0$$

$$J_k = \cancel{\frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} \lambda_j} + \sum_{j=k+1}^{d} \lambda_j = \sum_{j=k+1}^{d} \lambda_j$$

2.1a L_1 norm

l_2 norm



b. unsure, used answer key for hint

min $f(x)$ with the bound $\|x\|_p \leq k$

$$= \inf_{x \geq 0} \sup \mathcal{L}(x, \lambda) = \inf \sup_{\lambda \geq 0} f(x) + \lambda (\|x\|_p - k)$$

$$= f(x) + \lambda \|x\|_p$$

NMF

- fancy matrix decomposition such that gives us max information about X

$X = W \cdot H$

* X, W, H are non negative matrix
  - ↳ this is critical because ensuring its non negative keeps the data significant

$X \approx W \times H$ generally approximate

X represents non-neg matrix where each column is ~~[illegible]~~ observation, each row feature

reduces dimensionality

NMF assigns weights to the features to create a representative observation
- constrained to be positive

choose random W and H and update (constraints $\geq$ W, H ≥ 0)

1. coordinate descent : fix W, optimize H; fix H, optimize W

2. [insert multiplicative update] for each
   - ↳ each update finds a new local minimum

hw 7

c) $$\ell(\mu) = \sum_i \sum_k r_{ik} \log (\mathbb{P}(x_i \mid \theta_k))$$

$$= \sum_i \sum_k r_{ik} \sum_j x_{ij} \log \mu_{kj} + (1 - x_{ij}) \log (1 - \mu_{kj}))$$

done similar derivation in hw 6

maximum after taking partial derivative

$$\sum_i r_{ik} x_{ij} = \mu_{kj} \sum r_{ik}$$

b) $$\ell(\mu) = \sum_i \sum_k r_{ik} \log \mathbb{P}(x \mid \mu_k) + \log \mathbb{P}(\mu_k)$$

$$= \sum_i \sum_k r_{ik} \left(\sum x_{ij} \log \mu_{kj} + (1 - x_{ij}) \log (1 - \mu_{kj}) \right)$$

$$+ (a-1) \log \mu_{kj} + (b-1) \log (1 - \mu_{kj})$$

used a reverse ko simplifica

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\mu_{kj}(1 - \mu_{kj})} \cdot \sum_i r_{ik} x_{ij} - r_{ik} \mu_{kj} + a - 1 - \mu_{kj} a$$

$$+ \mu_{kj} - \mu_{kj} b + \mu_{kj}$$

$$= \sum r_{ik} x_{ij} - \sum r_{ik} x_{ij} - \left(\sum_i r_{ik} \, a + b - 2\right) \mu_{kj} + a - 1$$

$$\Rightarrow \sum_i r_{ik} x_{ij} + a - 1 = \mu_{kj} \sum r_{ik} + a + b - 2$$

hw 7

a) $l(\mu) = \sum_i \sum_k r_{ik} \log (P(x_i|\theta_k))$

$= \sum_i \sum_k r_{ik} \sum_j^d x_{ij} \log \mu_{kj} + (1-x_{ij}) \log (1-\mu_{kj})$

done similar derivation in hw6

maximum after taking partial derivative

$\sum_i r_{ik} x_{ij} = \mu_{kj} \sum r_{ik}$

b) $l(\mu) = \sum_i \sum_k r_{ik} \log P(x|\mu_k) + \log \Pi(\mu_k)$

$= \sum_i \sum_k r_{ik} \left(\sum x_{ij} \log \mu_{kj} + (1-x_{ij}) \log (1-\mu_{kj}) \right)$

$+ (a-1) \log \mu_{kj} + (b-1) \log (1-\mu_{k})$ used a few key simplification

$\frac{\partial l}{\partial \mu} = \frac{1}{\mu_{kj}(1-\mu_{kj})} \cdot \sum_i r_{ik} x_{ij} - r_{ik} \mu_{kj} + a - \mu_{kj} a +$

$+ \mu_{kj} - \mu_{kj} b + \mu_{kj}$

$= \sum r_{ik} x_{ij} - \sum r_{i} x_{ij} - \left(\sum_i r_{ik} + a + b - 2\right) \mu_{kj} + a - 1$

$\Rightarrow \sum_i r_{ik} x_{ij} + a - 1 = \mu_{kj} \sum r_{ik} + a + b - 2$

HW 6

$$\mu_K = \frac{\sum_i r_{ik} x_i}{r_k} \qquad \Sigma_k = \frac{1}{r_k} \sum_i r_{ik} x_i x^T - r_k \mu_k \mu^{-1}$$

$$l(\mu, \Sigma_k) = $$

$$= \frac{-1}{2} \sum r_{ik} (\log |\Sigma_k| + (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)$$

$$\frac{\partial}{\partial \mu_k} = r_{ik} \Sigma_k (x_i - \mu_K) = 0$$

$$\Rightarrow \sum_i r_{ik} x_i = \mu_k \sum_i r_{ik}$$

$$\Rightarrow \sum r_{ik} I = \left(\sum r_{ik} (x_i - \mu_k)(x_i - \mu_k)^T \right) \Sigma_k^{-1}$$