INFR10086 Machine Learning (MLG) Semester 1, 2022/23

Practice Exam

1. a) This is partially true, as sample complexity also depends on the confidence. Ignoring the factor of confidence, this statement is true, but the exact trade-off of samples and generalization error is hard to derive in most cases. We typically only have upper bounds and asymptotic bounds.

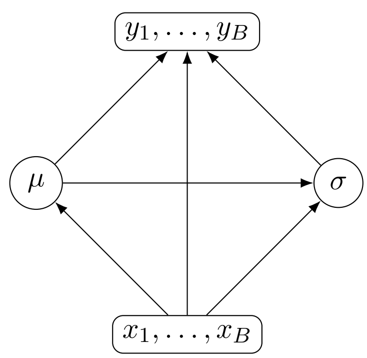
   b) VC dimension bounds the difference between training and generalization error, not test error. It is only a bound, meaning that the exact value can be smaller. Having a larger VC dimension does not mean the gap between training and generalization error is necessarily going to be larger.

   c) Test error is an approximation of the generalization error. We do not know what the generalization errors are.

   d) This statement depends on whether we can find another model with non-zero training error but a lower test error. We can only claim overfitting if we have successfully found such a model.

   e) Yes. With respect to ERM, any model is underfitting, so an overfitting model is likely also underfitting compared to ERM.

2. a) The computation graph is



   b)

$$\frac{\partial L}{\partial \sigma} = \sum_{i=1}^{B} \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial \sigma} = \sum_{i=1}^{B} \frac{\partial L}{\partial y_i} \frac{-(x_i - \mu)}{\sigma^2} \tag{1}$$

c)

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{B} \frac{\partial L}{\partial y_i}\frac{\partial y_i}{\partial \mu} + \frac{\partial L}{\partial \sigma}\frac{\partial \sigma}{\partial \mu} = \sum_{i=1}^{B} \frac{\partial L}{\partial y_i}\frac{-1}{\sigma} + \frac{\partial L}{\partial \sigma}\frac{1}{2}\frac{-2\mu}{\sqrt{\frac{1}{B}\sum_{i=1}^{B} x_i^2 - \mu^2}} \tag{2}$$

$$= \sum_{i=1}^{B} \frac{\partial L}{\partial y_i}\frac{-1}{\sigma} + \frac{\partial L}{\partial \sigma}\frac{-\mu}{\sigma} \tag{3}$$

d)

$$\frac{\partial L}{\partial x_b} = \sum_{i=1}^{B} \frac{\partial L}{\partial y_i}\frac{\partial y_i}{\partial x_b} + \frac{\partial L}{\partial \mu}\frac{\partial \mu}{\partial x_b} + \frac{\partial L}{\partial \sigma}\frac{\partial \sigma}{\partial x_b} \tag{4}$$

$$= \frac{\partial L}{\partial y_b}\frac{1}{\sigma} + \frac{\partial L}{\partial \mu}\frac{1}{B} + \frac{\partial L}{\partial \sigma}\frac{1}{2}\frac{2x_b/B}{\sqrt{\frac{1}{B}\sum_{i=1}^{B} x_i^2 - \mu^2}} \tag{5}$$

$$= \frac{\partial L}{\partial y_b}\frac{1}{\sigma} + \frac{\partial L}{\partial \mu}\frac{1}{B} + \frac{\partial L}{\partial \sigma}\frac{x_b}{B\sigma} \tag{6}$$

3. a)

$$L = \sum_{i=1}^{n} \mathbb{E}_{z\sim p(z|x_i)}[\log p(x_i|z)] + \mathbb{E}_{z\sim p(z|x_i)}\left[\log \frac{p(z)}{p(z|x_i)}\right] \tag{7}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{z\sim p(z|x_i)}\left[\log \frac{p(x_i|z)p(z)}{p(z|x_i)}\right] \tag{8}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{z\sim p(z|x_i)}\left[\log \frac{p(x_i, z)}{p(z|x_i)}\right] \tag{9}$$

$$= \sum_{i=1}^{n} \mathbb{E}_{z\sim p(z|x_i)}[\log p(x_i)] \tag{10}$$

$$= \sum_{i=1}^{n} \log p(x_i) \tag{11}$$

b) Once we have plugged in the distributions, we have

$$L = \sum_{i=1}^{n} \mathbb{E}_{z\sim p(z|x_i)}\left[-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_z|) - \frac{1}{2}(x-\mu_z)^\top \Sigma_z^{-1}(x-\mu_z)\right] \tag{12}$$

$$+ \mathbb{E}_{z\sim p(z|x_i)}\left[\frac{v_z}{p(z|x_i)}\right]. \tag{13}$$

We can see that L is a quadratic function of μ_z, and $\nabla^2_{\mu_z} L = -n\Sigma_z^{-1}$. The precision matrix Σ_z^{-1} is symmetric, so it is also positive semidefinite. This implies that L is concave in μ_z. To show that the precision matrix is symmetric,

$$I^\top = (\Sigma^{-1}\Sigma)^\top = \Sigma^\top(\Sigma^{-1})^\top = \Sigma(\Sigma^{-1})^\top, \tag{14}$$

so $(\Sigma^{-1})^\top = \Sigma^{-1}$.

c)

$$q(z|x) = p(z|x) = \frac{p(x,z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\sum_{z'} p(x|z')p(z')} \tag{15}$$

d) GMM becomes k-means if we choose

$$q(z|x) = \mathbb{1}_{z=\operatorname{argmin}_{k=1,\ldots,K} \|x-\mu_k\|}. \tag{16}$$