Homework 1: Introduction to Machine Learning, Spring 2018

Due: 2:00 pm, March 7 (Wednesday)

1. (10 pts) The multivariate Gaussian is the most widely-used joint distribution for continuous random variables. Suppose that $\boldsymbol{x} \in \mathbb{R}^D$ is a D-dimensional Gaussian random vector, i.e., $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix.

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

   Consider $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ where $\boldsymbol{A} \in \mathbb{R}^{d\times D}$ and $\boldsymbol{b} \in \mathbb{R}^d$. Show that $\boldsymbol{y}$ is also d-dimensional Gaussian and compute the mean and covariance of $\boldsymbol{y}$.

2. (10 pts) For any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, its **range space** Range($\boldsymbol{A}$) and **null space** Null($\boldsymbol{A}$) are defined as follows:

$$\begin{aligned} \text{Range}(\boldsymbol{A}) &= \{\boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \text{ for some } \boldsymbol{x} \in \mathbb{R}^n\} \\ \text{Null}(\boldsymbol{A}) &= \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = 0\}. \end{aligned}$$

   What are range space and null space of $\boldsymbol{A}$ and $\boldsymbol{B}$?

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

3. (10 pts) Consider two multivariate Gaussians $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where $\boldsymbol{x} \in \mathbb{R}^D$. Show that KL divergence $D_{KL}[p\|q]$ is calculated as:

$$D_{KL}[p\|q] = \frac{1}{2}\left[\text{tr}\left(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\right) + (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) - D + \log\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}\right].$$

4. (10 pts) The Jensen's inequality states that if $f(x)$ is a convex function and x is a random variable, then $\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$. Prove this when x is discrete random variable.

5. (10 pts) The Gibb's inequality states that $D_{KL}[p\|q] \geq 0$ with equality iff $p = q$. Prove this.