Tutorial Week 2

COMP9418 – Advanced Topics in Statistical Machine Learning, 17s2, UNSW Sydney

Instructor: Edwin V. Bonilla

Last Update: Monday 31st July, 2017 at 13:40

This tutorial provides a very small sample of problems you should be able to formalize and solve mathematically. If you struggle with the exercises below, I strongly advise you against taking COMP9418 for credits.

1. **Linear Algebra.** Given the matrix $\mathbf{A}$ and column vectors $\mathbf{x}$, $\mathbf{y}$:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \tag{1}$$

   (a) Compute $\mathbf{Ax}$, $\mathbf{x}^T\mathbf{y}$ and $\mathbf{xy}^T$

   (b) Find all the eigenvalues and eigenvectors of $\mathbf{A}$

2. **Expectation and variance.** Let $X \in \{0, 1\}$ be a Bernoulli random variable, i.e $p(x|\theta) = \theta^x(1-\theta)^{1-x}$. Derive expressions for the expectation $\mathbb{E}[X]$ and variance $\mathbb{V}[X]$. Show all your working.

3. **The Monty Hall problem.** You have entered a game where there are three boxes, only one of which contains a prize and the other two are empty. Your goal is to pick up the box with the prize in it. You select one of the boxes, and the host of the contest who knows the location of the prize and will not open up that box, opens one of the other boxes and reveals that it is empty. He then gives you to the chance to change your choice. Should you switch to another box? would that increase your chances of winning the prize?

4. **Unconstrained optimization.** Let $\mathbf{A}$ be a positive definite (PD) symmetric matrix and $\mathbf{x}, \mathbf{b}$ be column vectors. Find the minimum of the function:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} + c. \tag{2}$$

5. **Constrained optimization.** Let $\mathcal{D} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ be a set of categorical data where $\mathbf{x}$ is a K-dimensional vector encoded using one-hot encoding, i.e. $x_j \in \{0, 1\}$ and $\sum_{k=1}^K x_k = 1$. Assume that $\mathbf{x}$ follows a Categorical distribution, i.e. $p(\mathbf{x}) = \text{Cat}(\mathbf{x}|\boldsymbol{\theta}) = \prod_{k=1}^K \theta_k^{x_k}$, with $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, $\theta_k \geq 0$, and $\sum_k^K \theta_k = 1$.

   (a) Write down the likelihood of the observations given the model parameters $p(\mathcal{D}|\boldsymbol{\theta})$.

   (b) Find the maximum of the data log-likelihood $\mathcal{L} = \log p(\mathcal{D}|\boldsymbol{\theta})$ subject to the constraint $\sum_{k=1}^K \theta_k = 1$ and derive maximum likelihood estimates $\widehat{\theta}_{k\text{ML}}$. HINT: Use Lagrange multipliers.

6. **Conjugate priors.** Let $\text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}) \stackrel{\text{def}}{=} \frac{1}{\text{B}(\boldsymbol{\alpha})} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$ denote a Dirichlet distribution with hyperparameters $\boldsymbol{\alpha}$, where $\text{B}(\boldsymbol{\alpha})$ is the normalization constant given by the multivariate Beta function $\text{B}(\boldsymbol{\alpha}) = \int \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma(\alpha_0)}$; $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ is the Gamma function, which satisfies $\Gamma(1) = 1$ and $\Gamma(x+1) = x\Gamma(x)$; and $\alpha_0 = \sum_{k=1}^{K} \alpha_k$. Show that when using the likelihood $p(\mathcal{D}|\boldsymbol{\theta})$ in item 5a above and a Dirichlet prior $p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha})$, the posterior distribution is $p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\alpha}) = \text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha} + \mathbf{m})$, where $\mathbf{m} = (m_1, \ldots, m_K)$ and $m_k = \sum_{i=1}^{N} x_k^{(i)}$. Show all your working.