Deep Learning Quiz

Date: 9/19/2019. Duration: 50 minutes

1. Answer True or False on the following statements about the Gaussian distribution $\mathcal{N}$:
   (a) If x ~ $\mathcal{N}$, then x can be regarded as the sum of a large amount of underlying factors. (5%)
   (b) Out of all possible probability distributions (over real numbers) with the same variance, $\mathcal{N}$ encodes the maximum amount of uncertainty. (5%)
   (c) Any smooth density can be approximated by a mixture of Gaussian distributions with enough components. (5%)
2. Given a symmetric matrix $A \in \mathbb{R}^{d\times d}$. Show that for any $x \in \mathbb{R}^d$, the Rayleigh's Quotient
$$\lambda_{\min} \leq \frac{x^T A x}{x^T x} \leq \lambda_{\max}$$
holds, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of A. (15%)
3. Given a quadratic function $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x) = \frac{1}{2}x^{\mathrm{T}}Ax - b^{\mathrm{T}}x + c$, where $A \in \mathbb{R}^{2\times 2}$ is symmetric. Draw the graph of f when A is
   (a) positive definite; (5%)
   (b) positive semidefinite and singular. (5%)
4. Let $\mathrm{y} = f(\mathrm{x}; w) = w^{\mathrm{T}}\mathrm{x}$ be a random variable transformed from another multi-variate random variable x by a deterministic function f. Show that the variance of y, σ_{y}^2, equals $w^{\mathrm{T}}\Sigma_{\mathrm{x}}w$. (10%)
5. Consider a situation where a doctor wants to inference if a patient is having either the disease $\mathrm{y}^{(1)}$ or $\mathrm{y}^{(2)}$ by examining the patient's symptoms x. Explain why the Bayes' rule,
$$\mathrm{P}(\mathrm{y}|\mathrm{x}) = \frac{\mathrm{P}(\mathrm{x}|\mathrm{y})\mathrm{P}(\mathrm{y})}{\mathrm{P}(\mathrm{x})},$$
can make the inference easier. (10%)
6. Consider a random variable x ~ Uniform having 8 possible states with corresponding probabilities: $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$. What is the Shannon entropy of x, i.e., $\mathrm{H}(\mathrm{x}\sim\mathrm{P}) = -\sum_x P(x)\lg P(x)$? (10%)
7. When you implement the line $c = \log\left(\frac{a}{a+b}\right)$ for some a, $b > 0$ in a computer program which stores a, b and c as floats, what numerical issues may occur? (10%) How to walk around these issues in your implementation? (10%)
8. Consider a continuous, differentiable function $f: \mathbb{R}^d \to \mathbb{R}$ and an input point $a \in \mathbb{R}^d$.
   (a) For any direction u in the input space, show that the directional derivative of f at a along u equals to $\nabla f(a)^{\mathrm{T}}u$. (10% Hint: the directional derivative of f at a along u is the derivative of function $f(a + \varepsilon u)$ with respect to ε, evaluated at $\varepsilon = 0$.)
   (b) What is the direction in the input space that leads to the steepest decent of f starting from a, i.e., what is the solution of $\mathrm{argmin}_{u,||u||=1}\nabla f(a)^{\mathrm{T}}u$? (10%)