Final Exam of MML
09:10 – 12:00, 26 December 2023; Room E416
Note: Cell phones and any calculator are forbidden.

Part I: True (T) or False (F) (50%; each for 5%)

1. For any $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{A}^\top \mathbf{A}$ is invertible.
2. The shape of $\frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{x})\mathbf{x}$ is $\mathbb{R}^{1\times n}$ for $\mathbf{x} \in \mathbb{R}^n$.
3. Every positive definite matrix $M \in \mathbb{R}^{n\times n}$ is invertible.
4. For any objective convex function $f : \mathbb{R}^d \mapsto \mathbb{R}$, gradient descent with a fixed learning rate $\gamma \geq 0$ always converges to the global minimum of f.
5. $f : \mathbb{R}^2 \mapsto \mathbb{R}$; $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ is a convex function.
6. For any function ψ with arguments $\mathbf{x}, \mathbf{y}$, we have $\min_{\mathbf{y}} \max_{\mathbf{x}} \psi(\mathbf{x}, \mathbf{y}) \leq \max_{\mathbf{x}} \min_{\mathbf{y}} \psi(\mathbf{x}, \mathbf{y})$.
7. Given a set of N samples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ and probability densities $p(\mathbf{x} \mid \theta)$ parameterized by θ, the negative log-likelihood of the data is $-\prod_{i=1}^{N} \log p(y_i \mid \mathbf{x}_i, \theta)$.
8. Compared with the maximum a posteriori estimation (MAP), maximum likelihood estimation (MLE) suffers less overfitting issues.
9. Given a sampler of $\mathcal{N}(\mathbf{0}, \mathbf{I})$, we can generate a sample $\mathbf{y} \sim \mathcal{N}(\mu, \mathbf{I})$ by transforming $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ to $\mathbf{y} = \mathbf{x} + \mu$.
10. For $\mathbf{x} \sim \mathcal{N}(\mu_x, \Sigma_x)$ and $\mathbf{y} \sim \mathcal{N}(\mu_y, \Sigma_y)$, the distribution of $a\mathbf{x} + b\mathbf{y}$ for $a, b \in \mathbb{R}$ is given as $\mathcal{N}(a\mu_x + b\mu_y, a\Sigma_x + b\Sigma_y)$.

Part II: Calculations. (70%; each for 5%; ONLY THE ANSWERS ARE REQUIRED)

1. (5%) Compute $\frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{x})\mathbf{x}$.
2. (5%) For $\mu, \sigma \in \mathbb{R}$, compute the derivative $f'(x)$ of the function $f : \mathbb{R} \mapsto \mathbb{R}$;
$$f(x) = \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$
3. (5%) Let X be a continuous random variable with pdf $f_X : [0,1] \mapsto [0,1]$: $f_X(x) = 4x^3$. compute the pdf of $Y = X^2$.
4. (5%) Given $\mathbf{x}, \mathbf{y}, \mathbf{b} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{n\times n}$, if $\mathbf{x}, \mathbf{y}$ are random vectors such that $\mathbf{y} = 2\mathbf{A}\mathbf{x} - \mathbf{b}$ and $\mathbb{V}[\mathbf{x}] = \sigma$, then compute the variance $\mathbb{V}[\mathbf{y}]$. (Hint: $\mathbb{V}[\mathbf{x}] = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]^\top$)
5. (10%) Consider the problem:
$$\min_{\mathbf{x}\in\mathbb{R}^d} \mathbf{c}^\top \mathbf{x}$$
subject to $\mathbf{A}\mathbf{x} \preceq \mathbf{b}$, for $\mathbf{A} \in \mathbb{R}^{m\times d}$, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^d$.

   a. (5%) Please provide its Lagrangian function $\mathcal{L}(\mathbf{x}, \lambda)$.

   b. (5%) Please list the dual problem.

6. (5%) Given $f : \mathbb{R}^+ \mapsto \mathbb{R}$; $f(x) = x \ln x$. Compute $f(t) + (\nabla_x f)(t)^\top (z - t)$ for $t = e^2, z = e^3$.

7. (10%) Given Billingsley's Theorem: Let $f_X(\mathbf{x})$ be the pdf of the multivariate continuous random variable X. If the function $\mathbf{y} = U(\mathbf{x})$ is differentiable and invertible for all values within the domain of $\mathbf{x}$, then for corresponding values of $\mathbf{y}$, the pdf of $Y = U(X)$ is given by

$$f(\mathbf{y}) = f_{\mathbf{x}}(U^{-1}(\mathbf{y})) \cdot \left| \det \left(\frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right) \right| .$$

Consider a bivariate random variable X with states $\mathbf{x} = [x_1 \ x_2]^\top$, the pdf $f(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} \mathbf{x}^\top \mathbf{x} \right)$ and a matrix $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. Find the pdf of the random variable $\mathbf{y} = \mathbf{A}\mathbf{x}$.

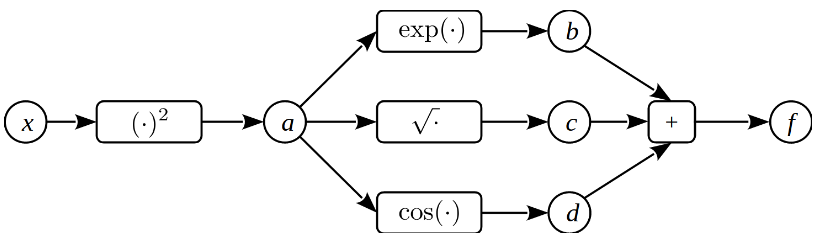
8. (10%) Consider $f : \mathbb{R}^2 \mapsto \mathbb{R};\ f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{x} - \begin{bmatrix} 2 \\ 3 \end{bmatrix}^\top \mathbf{x} + [1 \ \ 0]^\top$.

   a. (5%) Compute $\nabla f(\mathbf{x})$.

   b. (5%) Set the step size $\gamma = 0.1$ and initial $\mathbf{x}_0 = [0 \ \ 0]^\top$, compute the iterative solution $\mathbf{x}_2$ given by the gradient descent algorithm.

9. (5%) Given the function $f(x, y) = x^2 + 2xy$ for $x, y \in \mathbb{R}$, please compute the Hessian matrix of f.

10. (10%) Consider the following workflow



According to the automatic differentiation rule of the reverse mode (backpropagation), please write down how we can compute $\frac{\partial f}{\partial a}$ and $\frac{\partial f}{\partial x}$.

(Hint: $\frac{\partial f}{\partial b} = \frac{\partial f}{\partial c} = \frac{\partial f}{\partial d} = 1$)