Calculus: Homework 4

1. (1.5 pt) Calculate the second-degree Taylor polynomial of

$$f(x,y) = e^{(x^2+y^2)}$$

at the point $(0,0)$.

2. (1.5 pt) Find the gradient of the following functions at the given point p

- $f(x,y) = (ax^2 - x^2 - y^2)^{1/2}$ at $p = (a/2, a/2)$
- $g(x,y) = e^y \cos(3x+y)$ at $p = (\frac{2\pi}{3}, 0)$
- $h(x_1, \ldots, x_n) = \sum_{i=1}^n x_1 x_i$ at $p = (1, \ldots, 1)$

3. (1.5 pt) Find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ for the following functions

- $x \cos(x) \sin(y)$
- $(x^2+y^2)\ln(x^2+y^2)$

4. (1.5 pt) For all $1 \leq j \leq n$ find the partial derivative $\frac{\partial f}{\partial x_j}$ of the following functions

- $f(x_1, \ldots, x_n) = \Pi_{i=1}^n x_i^i$
- $g(x_1, \ldots, x_n) = \sum_{i=2}^n (x_{i-1} + x_i)$

5. (2 pt) A *multivariate Gaussian* is a probability distribution of the form

$$p(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x-u)^T \Sigma^{-1}(x-u)}$$

where Σ is a symmetric $n \times n$ matrix, $u = (u_1, \ldots, u_n)$ is a vector of size n and $\vec{x} = (x_1, \ldots, x_n)$. As usual, we denote by A^T the transpose of a matrix A and $\det(\Sigma)$ is the determinant of Σ. Note that

$$-\frac{1}{2}(x-u)^T \Sigma^{-1}(x-u)$$

is a scalar. Compute

$$\frac{\partial p}{\partial x_1}$$

in the case where Σ is a diagonal matrix $\mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$:

$$\begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

Hint: it is helpful to first rewrite p such that it does not contain any matrix operation anymore. First prove that

$$-\frac{1}{2}(x-u)^T \Sigma^{-1}(x-u) = -\frac{1}{2}\sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i - u_i)^2$$

The expression $(x-u)^T\Sigma^{-1}(x-u)$ is a nice and compact way to write the right-side sum, but it makes it difficult to compute the partial derivative. This is the reason why we first develop the matrix product. After doing this, calculate the partial derivative in the usual way.

6. (1 pt) We consider a probability distribution $p(x_1,\ldots,x_n)$ defined over the set $\{(x_1,\ldots,x_n) : x_1 > 0,\ldots,x_n > 0\}$ by

$$p(x_1,\ldots,x_n) = \Pi_{i=1}^{n} p_i(x_i)$$

with

$$p_i(x_i) = \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i}$$

where α and β are constants and Γ is a function from $\mathbb{R}$ to $\mathbb{R}$, which you can regard as a constant as well when calculating the partial derivative (for your information, $p_i(x_i)$ is a so-called Gamma distribution).

For all $1 \le j \le n$, find

$$\frac{\partial p}{\partial x_j}$$

7. (1 pt) In neural networks, the *softmax function* is often used as the last layer. This function gets as input a vector of real numbers and rescales them so that the resulting vector sums to 1 (and can be interpreted as a probability distribution).

More formally, suppose you are given n numbers: $\vec{z} = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^n$. The softmax function then calculates for each $j \in \{1,\ldots,n\}$:

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{n} e^{z_k}}$$

Calculate the partial derivatives of this softmax function with respect to the data points z_i. That is, compute for $1 \le i \le n$:

$$\frac{\partial p_j}{\partial z_i}$$

Hint: consider the case where $i = j$ separately from the case where $i \neq j$.