HOMEWORK 6
NO DUE DATE

Do the following exercises without turning them in. There is no credit for this assignment, but doing these problems will help you prepare for the end sem exam.

Problem 1. Recall from linear algebra that $||\cdot||$ is a vector norm if (a) $||x|| > 0$ for $x \neq 0$, (b) $||\alpha x|| = |\alpha|\,||x||$ and (c) $||x + y|| \leq ||x|| + ||y||$. The most commonly used example of a vector norm is the p-norm, which is defined by

$$||x||_p = \left(\sum_{n=1}^{N} |x|^p \right)^{\frac{1}{p}} \tag{1}$$

when $p \geq 1$.

a. Show that every vector norm (not just the p-norm) is a convex function.

b. Show that for fixed $t \geq 0$ the function

$$f(w) = ||Xw - y||_2^2 + t||w||$$

is convex. Here the first norm is the 2-norm and the second norm can be any norm (e.g., in Lasso the second norm is the 1-norm). Moreover, $X \in \mathbb{R}^{N \times D}$, $w \in \mathbb{R}^D$, $y \in \mathbb{R}^N$. Be careful about the square in the 2-norm.

c. Give an example of two vectors that shows that (1) does not satisfy (c) if $p < 1$. Hence this case does does not define a norm.

d. Draw a picture that shows the vectors in $\mathbb{R}^2$ with the property that $||x||_p = 1$, for $p = 1, 2, 4$. Using this picture, explain why the maximum norm is also referred to as the ∞-norm.

e. Draw a similar picture that shows the vectors in $\mathbb{R}^2$ with the property that $||x||_p = 1$, for $p = 1, 1/2, 1/4$. Explain what is meant by $||x||_0$.

Problem 2. When we derived the MLE at the beginning of Chapter 11 we had a single output variable. Suppose we have multiple outputs with normally distributed uncertainties, i.e.,

$$p(Y = y_k | X = x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y_k - w_k \cdot x)^2}{2\sigma_k^2} \right), \quad k = 1, \ldots, C. \tag{2}$$

Here, $x \in \mathbb{R}^D$, and the parameters w_k are vectors in $\mathbb{R}^D$ which are the columns of the matrix $W \in \mathbb{R}^{D \times C}$. Derive the MLE in this case and derive the form of the corresponding least squares problem. Derive the normal equation by setting the partial derivatives with respect to the $w_{k,d}$ to zero.

Problem 3. For the nodes $0 = x_0 < x_1 < \cdots < x_N = 1$, the hat function is defined as

$$B_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & x_{i-1} \leq x \leq x_i, \\ \frac{x_i - x}{x_i - x_{i-1}} & x_i \leq x \leq x_{i+1}, \\ 0 & \text{else.} \end{cases}$$

The piecewise linear interpolation of a function $f(x)$ is given by

$$f_{lin}(x) = \sum_{i=0}^{N} f(x_i)B_i(x) \tag{3}$$

- **a.** Draw the graph of a few B_i's and draw the graph for some function $f(x)$ and its linear interpolation $f_{lin}(x)$.
- **b.** Express the evaluation of (3) as a neural network with one hidden layer, similar to what we did for piecewise constant splines in the lecture, page 13p5. Use ReLU as the activation function. Show how to compute the parameters in the neural net by direct calculation when the nodes x_i and the function values $f(x_i)$ are given.

Problem 4. The numerical cost (number of floating point operations) of evaluating the matrix vector product $y = Ax$ is $2nm$, where n is the number of columns and m is the number of rows. Likewise, when we use the obvious algorithm, the cost of the matrix matrix product AB is $2mnp$ where $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{p\times m}$.

Compute the cost of calculating the Jacobian of $f = f_L \circ .. \circ f_1$ if

- **a.** J_f is computed by multiplying out the product $J_L \dots J_1$ in the order $V_2 = J_2 J_1$, then $V_3 = J_3 V_2$ and so on.
- **b.** J_f is computed by multiplying out the product $J_L \dots J_1$ in the order $V_{L-1} = J_L J_{L-1}$, then $V_{L-2} = V_{L-1} J_{-2}$ and so on.
- **c.** J_f is computed by forward mode differentiation.
- **d.** J_f is computed by reverse mode differentiation.

See lecture notes, 13p8. Do not include the cost of setting up J_ℓ.

Problem 5. φ is an activation function with known derivative φ', and a, b, c, d, e, f and p, q, r are parameters. Moreover,

$$\begin{aligned} u &= \varphi(ax + b) \\ z_1 &= \varphi(cu + d) \\ z_2 &= \varphi(eu + f) \\ y &= pu + qz_1 + rz_2 \end{aligned}$$

- **a.** Draw a graph that shows the evaluation order for $y(x)$
- **b.** Compute the derivative $\frac{\partial y}{\partial a}$