TD 2

Exercice 1 – Reminders on Convexity

Q 1.1 Among the following functions, which are convex :
$f(x) = x\cos(x)$, $g(x) = -\log(x) + x^2$, $h(x) = x\sqrt{x}$, $t(x) = -\log(x) - \log(10 - x)$?

Q 1.2 Let $f \in \mathbb{R}^n \to \mathbb{R}$ be a linear map. Recall what the gradient of f is : $\nabla f(\mathbf{x})$. Give the gradient of $f(\mathbf{x}) = 2x_1 + x_2^2 + x_2x_3$.

Q 1.3 Express $\nabla_{\mathbf{x}}(f(\mathbf{x}) + g(\mathbf{x}))$, $\nabla_{\mathbf{x}} tf(\mathbf{x})$.
Give the expression for $\nabla_{\mathbf{x}} b'\mathbf{x}$ where $b \in \mathbb{R}^d$ and $\nabla_{\mathbf{x}}\mathbf{x}'A\mathbf{x}$ for symmetric A.

Exercice 2 – Linear Regression

Let a training data set $\mathcal{D} = \{\mathbf{x^i}, y^i\}_{i=1,\ldots,N}$, where $\mathbf{x^i} \in \mathbb{R}^d$ and $y^i \in \mathbb{R}$.
By convention, which will be followed throughout the course, the data matrix will be denoted by X, where each row corresponds to one example. The response matrix Y is thus a column matrix, and the weight matrix W as well. The error on $\mathcal{D}$ will be denoted by $C(W)$.

Q 2.1 Analytical Solution

Q 2.1.1 Recall the principle of linear regression. Which error function $C(W)$ is used ?

Q 2.1.2 What are the dimensions of the matrices X, W, and Y ? Recall the matrix formulation of the error.

Q 2.1.3 Find analytically the matrix W solution of the linear regression that minimizes $C(W)$.

Q 2.1.4 The same question if we now consider a linear machine with a bias. What is the optimal value of the bias w_0 in this case ?

Q 2.2 Recall the principle of the gradient descent algorithm. Apply it to the case of linear regression.

Q 2.3 Consider a 2-dimensional problem.

Q 2.3.1 Plot the parameter space in 2D. Arbitrarily position the points $\mathbf{w}^0$, the initial point, and $\mathbf{w}^*$, the analytical solution of the problem. Given the quadratic nature of the cost, draw the iso-contours of the cost function in the parameter space. What is the shape of the cost function $C(\mathbf{w}^0)$ in the parameter space ?

Q 2.3.2 Draw the vector $\nabla C(\mathbf{w}^0)$. What does this vector represent geometrically ?

Exercice 3 – Logistic Regression

Q 3.1 Consider a data set X with binary labels $Y = \{-1, 1\}$. In logistic regression, it is assumed that the log-odds of the conditional probabilities $p(y|\mathbf{x})$ can be modeled by a linear function :

$$\log\left(\frac{p(y=1|\mathbf{x})}{1-p(y=1|\mathbf{x})}\right) = \mathbf{w}\cdot\mathbf{x}.$$

- What is the goal ?
- What label should be predicted for $\mathbf{x}$ if $\mathbf{w}\cdot\mathbf{x} > 0$?
- What is $p(y = 1|\mathbf{x})$? Plot the function $p(y = 1|\mathbf{x})$ as a function of $\mathbf{w}\cdot\mathbf{x}$.

- Deduce the type of boundary that logistic regression can determine.
- Provide the expression for $p(y = -1|\mathbf{x})$ and then for $p(y = y_i|\mathbf{x})$ in a compact form.

Q 3.2 For a single dimension x_i, what is the influence of its value on $p(y = 1|\mathbf{x})$? What is the limit of logistic regression ?

Q 3.3 What is the expression for the conditional likelihood of $\mathbf{w}$ given an example $(\mathbf{x}, y)$? The log-likelihood ? And for a set of examples $\mathcal{D}$?

Q 3.4 Propose an algorithm to solve the logistic regression problem.