# ETC3555 2018 - Lab 4

Linear models and gradient descent (I/II)

*Cameron Roach and Souhaib Ben Taieb*

*15 August 2018*

**Exercise 1**

Solve exercise 3.7 in Learning From Data.

**Exercise 2**

Solve exercise 3.9 in Learning From Data.

**Exercise 3**

Solve problem 3.16 in Learning From Data.

## Exercise 3.7

For logistic regression, show that

$$\nabla E_{\text{in}}(\mathbf{w}) = -\frac{1}{N}\sum_{n=1}^{N}\frac{y_n\mathbf{x}_n}{1+e^{y_n\mathbf{w}^{\mathrm{T}}\mathbf{x}_n}}$$
$$= \frac{1}{N}\sum_{n=1}^{N} -y_n\mathbf{x}_n\theta(-y_n\mathbf{w}^{\mathrm{T}}\mathbf{x}_n).$$

Argue that a 'misclassified' example contributes more to the gradient than a correctly classified one.

Figure 1: Source: Abu-Mostafa et al. Learning from data. AMLbook.

## Exercise 3.9

Consider pointwise error measures $\mathrm{e}_{\text{class}}(s,y) = [\![y \neq \text{sign}(s)]\!]$, $\mathrm{e}_{\text{sq}}(s,y) = (y-s)^2$, and $\mathrm{e}_{\log}(s,y) = \ln(1+\exp(-ys))$, where the signal $s = \mathbf{w}^{\mathrm{T}}\mathbf{x}$.

(a) For $y = +1$, plot $\mathrm{e}_{\text{class}}$, $\mathrm{e}_{\text{sq}}$ and $\frac{1}{\ln 2}\mathrm{e}_{\log}$ versus $s$, on the same plot.

(b) Show that $\mathrm{e}_{\text{class}}(s,y) \leq \mathrm{e}_{\text{sq}}(s,y)$, and hence that the classification error is upper bounded by the squared error.

(c) Show that $\mathrm{e}_{\text{class}}(s,y) \leq \frac{1}{\ln 2}\mathrm{e}_{\log}(s,y)$, and, as in part (b), get an upper bound (up to a constant factor) using the logistic regression error.

These bounds indicate that minimizing the squared or logistic regression error should also decrease the classification error, which justifies using the weights returned by linear or logistic regression as approximations for classification.

Figure 2: Source: Abu-Mostafa et al. Learning from data. AMLbook.

**Problem 3.16** In Example 3.4, it is mentioned that the output of the final hypothesis $g(\mathbf{x})$ learned using logistic regression can be thresholded to get a 'hard' ($\pm 1$) classification. This problem shows how to use the risk matrix introduced in Example 1.1 to obtain such a threshold.

Consider fingerprint verification, as in Example 1.1. After learning from the data using logistic regression, you produce the final hypothesis

$$g(\mathbf{x}) = \mathbb{P}[y = +1 \mid \mathbf{x}],$$

which is your estimate of the probability that $y = +1$. Suppose that the cost matrix is given by

| | | True classification | |
|---|---|---|---|
| | | $+1$ (correct person) | $-1$ (intruder) |
| you say | $+1$ | 0 | $c_a$ |
| | $-1$ | $c_r$ | 0 |

For a new person with fingerprint $\mathbf{x}$, you compute $g(\mathbf{x})$ and you now need to decide whether to accept or reject the person (i.e., you need a hard classification). So, you will accept if $g(\mathbf{x}) \geq \kappa$, where $\kappa$ is the threshold.

(a) Define the cost(accept) as your expected cost if you accept the person. Similarly define cost(reject). Show that

$$\begin{aligned} \text{cost(accept)} &= (1 - g(\mathbf{x}))c_a, \\ \text{cost(reject)} &= g(\mathbf{x})c_r. \end{aligned}$$

(b) Use part (a) to derive a condition on $g(\mathbf{x})$ for accepting the person and hence show that

$$\kappa = \frac{c_a}{c_a + c_r}.$$

(c) Use the cost-matrices for the Supermarket and CIA applications in Example 3.4 to compute the threshold $\kappa$ for each of these two cases. Give some intuition for the thresholds you get.

Figure 3: Source: Abu-Mostafa et al. Learning from data. AMLbook.