Statistical Machine Learning 2016

Exercises, week 1

Exercise 1

In analyzing problems in which a sigma-summation symbol is involved, it is sometimes helpful to write out the sum. By writing out the sum, I mean e.g.,

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

or more general

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n \ .$$

- Show, by explicitly writing out the sums, rearranging terms, and using brackets where needed, that the following four equations hold:

$$\sum_{i=1}^{3}(ax_i) = a\big(\sum_{i=1}^{3} x_i\big) \tag{1}$$

$$\sum_{i=1}^{3}\big(\sum_{j=1}^{2} a_{ij}\big) = \sum_{j=1}^{2}\big(\sum_{i=1}^{3} a_{ij}\big) \tag{2}$$

$$\sum_{i=1}^{3}\big(\sum_{j=1}^{2} x_i y_j\big) = \big(\sum_{i=1}^{3} x_i\big)\big(\sum_{j=1}^{2} y_j\big) \tag{3}$$

$$\sum_{i=1}^{3} a = 3a \tag{4}$$

Exercise 2

Calculate the gradient ∇f of the following functions $f(\mathbf{x})$. In the left column, $\mathbf{x} = (x_1, x_2, x_3)$. In the right column, $\mathbf{x} = (x_1, \ldots, x_n)$.

a) $f(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3$	e) $f(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i$
b) $f(x_1, x_2, x_3) = x_2$	f) $f(\mathbf{x}) = x_i$
c) $f(x_1, x_2, x_3) = x_1x_2x_3$	g) $f(\mathbf{x}) = \prod_{i=1}^{n} x_i$
d) $f(x_1, x_2, x_3) = x_1^{k_1}x_2^{k_2}x_3^{k_3}$	h) $f(\mathbf{x}) = \prod_{i=1}^{n} x_i^{k_i}$

Note: often it suffices to write down the partial derivative $\partial f/\partial x_j$ (Can you tell why?).

Exercise 3

The function

$$f(x,y) = 2x^2 - xy + y^2 - x + y + 5.5 \tag{5}$$

has a unique minimum (x^*, y^*). Calculate this point.

Exercise 4

Calculate the minimum x^* of the following two functions.

$$f(x) = \sum_{i=1}^{n} (x - a_i)^2 \tag{6}$$

$$f(x) = \sum_{i=1}^{n} \alpha_i (x - a_i)^2 \quad (\text{with } \alpha_i > 0) \tag{7}$$

Exercise 5

Calculate the gradient ∇f of

$$f(\vec{h}) = \sum_{i=1}^{n} p_i h_i - \ln\Big(\sum_{i=1}^{n} \exp(h_i)\Big) \tag{8}$$

Exercise 6

Compute the minimum x^* of

$$f(x) = a\ln(x) + \frac{b}{2x^2} \tag{9}$$

with $a > 0$, $b > 0$ en $x > 0$. Express your answer in terms of a en b. (Note: $\ln(x)' = 1/x$).

Exercise 7

(see Bishop, appendix C, eq.C.1) An $N \times M$ matrix $\mathbf{A}$ has elements A_{ij} (with i the row- and j the columnindex). The transposed matrix $\mathbf{A}^T$ has elements $(\mathbf{A}^T)_{ij} = A_{ji}$. By writing out the matrix product using index notation show that

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T \tag{10}$$

Hint: $\mathbf{C} = \mathbf{AB}$ corresponds to $C_{ij} = \sum_{k=1}^{M} A_{ik}B_{kj}$

Exercise 8

(Exercise 1.1 from the Bishop book.) Consider the M-th order polynomial

$$y(x;\mathbf{w}) = w_0 + w_1 x + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j \tag{11}$$

and the error function

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} \{y(x_n;\mathbf{w}) - t_n\}^2 \tag{12}$$

with x_n, t_n the input/output pairs from the data set. Define the error per data point as

$$E_n(\mathbf{w}) = \frac{1}{2}\{y(x_n; \mathbf{w}) - t_n\}^2 \tag{13}$$

(so $E = \sum_{n=1}^{N} E_n$). Note that $x =$ 1-dimensional, and that in this exercise the super-indices i, j represent 'power'.

1. Calculate the gradient of the error per data point E_n:

$$\nabla E_n \quad (= (\frac{\partial E_n}{\partial w_0}, \ldots, \frac{\partial E_n}{\partial w_M})^T\ ). \tag{14}$$

2. Calculate the gradient of the total error E.

3. Show that the partial derivatives can be written as

$$\frac{\partial E}{\partial w_i} = \sum_{j=0}^{M} A_{ij} w_j - T_i \tag{15}$$

with A_{ij} and T_i defined as

$$A_{ij} = \sum_{n=1}^{N} x_n^{i+j} \qquad T_i = \sum_{n=1}^{N} t_n x_n^i. \tag{16}$$

4. When E is minimal it holds that $\nabla E = 0$ (i.e., all partial derivatives are zero). Using this, show that in the minimum of E the parameters $\mathbf{w}$ satisfy

$$\sum_{j=0}^{M} A_{ij} w_j = T_i. \tag{17}$$