Calculus summary handout 1

Wim Wiegerinck

June 1, 2005

1 Logarithm and exponential function

Two important functions in statistical pattern recognition are the exponential function and the logarithm (they are actually each others inverses).

1.1 Exponential function

Some roperties:

$$\begin{aligned}
\exp(x)\exp(y) &= \exp(x+y) \\
(\exp(x))^a &= \exp(ax) \\
1/\exp(x) &= \exp(-x) \\
\exp(0) &= 1 \\
\frac{d}{dx}\exp(x) &= \exp(x)
\end{aligned}$$

In addition, it has the following limits,

$$\lim_{x\to-\infty}\exp(x) = 0 \tag{1}$$

$$\lim_{x\to\infty}\exp(x) = \infty \tag{2}$$

where it should be noted that the exponential function is a very rapidly increasing function: for any power n (no matter how large) $x^n/\exp(x) \to 0$ if $x \to \infty$. In words: 'the exponential function wins from any power'.

1.2 Logarithm

The logarithm is written as $\log(x)$ or $\ln(x)$. Unless stated otherwise, the logarithm is to the base of e (or: natural logarithm), even if it is written as $\log(x)$. The logarithm is only defined for positive x (we will not deal with complex numbers).

Properties:

$$\log(xy) = \log(x) + \log(y) \tag{3}$$

$$\log(x/y) = \log(x) - \log(y) \tag{4}$$

$$\log(x^a) = a\log(x) \tag{5}$$

$$\log(1/x) = -\log(x) \tag{6}$$

$$\log(1) = 0 \tag{7}$$

$$\frac{d}{dx}\log(x) = \frac{1}{x} \tag{8}$$

In addition, it has the following limits,

$$\lim_{x\downarrow 0} \log(x) = -\infty \tag{9}$$

$$\lim_{x\to\infty} \log(x) = \infty \tag{10}$$

where it should be noted that the logarithmic function is a very slowly increasing function: for any power $\alpha > 0$ (no matter how small), $\log(x)/x^\alpha \to 0$ if $x \to \infty$. In words: 'any power wins from the logarithmic function'.

The exp and the log are each others inverses, so

$$\log(\exp(x)) = x \tag{11}$$

$$\exp(\log(x)) = x \quad (\text{for } \ x > 0) \tag{12}$$

2 Partial derivatives, gradient, nabla symbol

Let $f(x_1, \ldots, x_n) = f(\mathbf{x})$ be a function of several variables. The gradient of f, denoted as ∇f (the symbol " ∇" is called 'nabla') , is the vector of all partial derivatives:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^T$$

NB, the partial derivative $\partial f(x_1, \ldots, x_i, \ldots, x_n)/\partial x_i$ is computed by taking the derivative with respect to x_i while keeping all other variables constant.

Example:

$$f(x, y, z) = xy^2 + 3.1yz$$

Then

$$\nabla f(x, y, z) = \left(y^2, 2xy + 3.1z, 3.1y\right)^T$$

At local minima (and maxima, and so-called saddle points) of a differentiable function f, the gradient is zero, i.e., $\nabla f = 0$.

Example:

$$f(x, y) = x^2 + y^2 + (y + 1)x$$

So

$$\nabla f(x, y) = (2x + y + 1, 2y + x)^T$$

Then we can compute the point (x^*, y^*) that minimizes f by setting $\nabla f = 0$,

$$\left.\begin{array}{ll} 2x^* + y^* + 1 & = 0 \\ 2y^* + x^* & = 0 \end{array}\right\} \Rightarrow (x^*, y^*) = (-\frac{2}{3}, \frac{1}{3}) \tag{13}$$

2.1 Chain rule

Suppose f is a function of $y_1, y_2, ..., y_k$ and each y_j is a function of x, then we can compute the derivative of f with respect to x by the chain rule

$$\frac{df}{dx} = \sum_{j=1}^{k} \frac{\partial f}{\partial y_j} \frac{dy_j}{\partial x}$$

Example:

$$f(y(x), z(x)) = y(x)/z(x)$$

and $y(x) = x^4$ and $z = x^2$ then

$$\frac{df}{dx} = \frac{1}{z(x)} y'(x) - \frac{y(x)}{z(x)^2} z'(x) = 2x$$

3 Taylor series / Taylor expansions

3.1 Taylor expansion in one dimension

Assuming that $f(x)$ has derivatives of all orders in $x = a$, then the Taylor expansion of f around a is

$$f(a+\epsilon) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}\epsilon^k = f(a) + \epsilon f^{'}(a) + \frac{\epsilon^2}{2} f^{''}(a) + \dots$$

The prefactors in the Taylor series can be checked by computing the Taylor expansion of a polynomial.

Linearization of a function around a is taking the Taylor expansion up to first order:

$$f(a+x) = f(a) + xf'(a)$$

Examples: check that for small x the following expansions are correct up to second order:

$$\begin{aligned}
\sin(x) &= x \\
\cos(x) &= 1 - \frac{1}{2}x^2 \\
\exp(x) &= 1 + x + \frac{1}{2}x^2 \\
(1+x)^a &= 1 + ax + \frac{a(a-1)}{2}x^2 \\
\ln(1+x) &= x - \frac{1}{2}x^2
\end{aligned}$$

3.2 Taylor expansion in several dimensions

The Taylor expansion of a function of several variables, $f(x_1, \dots, x_n) = f(\mathbf{x})$ is (up to second order)

$$f(\mathbf{a}+\epsilon) = f(\mathbf{a}) + \epsilon^T \nabla f(\mathbf{a}) + \frac{1}{2}\epsilon^T H \epsilon$$

with H the Hessian, which is the symmetric matrix of partial derivatives

$$H_{ij} = \left.\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})\right|_{\mathbf{x}=\mathbf{a}}$$

4 Summations and Kronecker delta

If $\mathbf{A}$ is a $N \times M$ matrix with entries A_{ij} and $\mathbf{v}$ an M-dimensional vector with entries v_i, then $\mathbf{w} = \mathbf{Av}$ is a N-dimensionsional vector with entries

$$w_i = \sum_{j=1}^{M} A_{ij} v_j \tag{14}$$

Similarly, if $\mathbf{B}$ is a $M \times K$ matrix with entries B_{ij}, then $\mathbf{C} = \mathbf{AB}$ is a $N \times K$ matrix with entries

$$C_{ik} = \sum_{j=1}^{M} A_{ij} B_{jk} \tag{15}$$

The notation δ_{ij} denotes usually the Kronecker delta symbol, i.e.,

$$\begin{cases} \delta_{ij} & = 1 \quad \text{if} \qquad\qquad i = j \\ \delta_{ij} & = 0 \quad \text{otherwise} \end{cases} \tag{16}$$

It has the nice property that it 'eats' dummy indices in summations:

$$\sum_{j=1}^{M} \delta_{ij} v_j = v_i \qquad \text{for all} \quad 1 \leq i \leq M \tag{17}$$

(The Kronecker delta can be viewed as the entries of the identity matrix $\mathbf{I}$. In vector notation, (17) is equivalent to the statement $\mathbf{Iv} = \mathbf{v}$.)

5 Dirac's delta function

The Dirac delta function $\delta(x)$ is defined such that

$$\delta(x) = 0 \;\; \text{if} \;\; x \neq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(x) \mathrm{d}x = 1$$

(actually, it is not a proper function, but that is another story). $\delta(x)$ is a spike (a peak) at $x = 0$. The function $\delta(x - x_0)$ as a function of x is a spike at x_0. As a consequence of the definition, the delta function has important property

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0) = f(x_0)$$

see preface xiii.