Machine Learning Module

Week 1

Tutorial, Week 1

Simple Matrix & Vector Differentiation for Least-Squares

Mark Girolami
girolami@dcs.gla.ac.uk
Department of Computing Science
University of Glasgow

January 8, 2006

1 Least-Squares Error Criterion in Matrix Notation

From the definition of the vectors $\mathbf{t}$, $\mathbf{w}$ and matrix $\mathbf{X}$ then we can obtain the matrix representation for the MSE as follows.

Define the nth element of $\mathbf{t}$ as t_n, the dth element of $\mathbf{w}$ as w_d and the element in the nth row and dth column of $\mathbf{X}$ as x_{nd}. Also define the nth row of $\mathbf{X}$ as $\mathbf{X}_n$ then the following steps will get us our desired expression.

$$\begin{aligned}
MSE &= \frac{1}{N}\sum_{n=1}^{N}(t_n - \sum_{d=1}^{D} w_d x_{nd})^2 & (1)\\
&= \frac{1}{N}\sum_{n=1}^{N}\left(t_n^2 - 2t_n\sum_{d=1}^{D} w_d x_{nd} + \left(\sum_{d=1}^{D} w_d x_{nd}\right)^2\right) & (2)\\
&= \frac{1}{N}\sum_{n=1}^{N} t_n t_n - \frac{2}{N}\sum_{n=1}^{N}\sum_{d=1}^{D} t_n w_d x_{nd} + \frac{1}{N}\sum_{n=1}^{N}\sum_{d=1}^{D}\sum_{d'=1}^{D} w_d x_{nd} w_{d'} x_{nd'} & (3)\\
&= \frac{1}{N}\mathbf{t}^\mathsf{T}\mathbf{t} - \frac{2}{N}\sum_{n=1}^{N} t_n \mathbf{X}_n \mathbf{w} + \frac{1}{N}\sum_{n=1}^{N}\mathbf{w}^\mathsf{T}\mathbf{X}_n^\mathsf{T}\mathbf{X}_n\mathbf{w} & (4)\\
&= \frac{1}{N}\mathbf{t}^\mathsf{T}\mathbf{t} - \frac{2}{N}\mathbf{t}^\mathsf{T}\mathbf{X}\mathbf{w} + \frac{1}{N}\mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w} & (5)
\end{aligned}$$

which follows from expansion of the quadratic form

$$MSE = \frac{1}{N}(\mathbf{t} - \mathbf{X}\mathbf{w})^\mathsf{T}(\mathbf{t} - \mathbf{X}\mathbf{w}) \tag{6}$$

2 Vector of Partial Derivatives of MSE

We will make use of The Matrix Cookbook to obtain the required derivatives directly in vector form. From Eqn (5) we see that the only components of MSE which are functions of $\mathbf{w}$ are the second and third. Now as the scalar form $\mathbf{a}^\mathsf{T}\mathbf{b} = \mathbf{b}^\mathsf{T}\mathbf{a}$ we know that

$$\mathbf{t}^\mathsf{T}\mathbf{X}\mathbf{w} = \mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{t} \tag{7}$$

and from Page 9 of The Matrix Cookbook (2.3.1) the derivative of a scalar inner product form follows as

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^\mathsf{T}\mathbf{b} = \mathbf{b} \tag{8}$$

and as $\mathbf{X}^\mathsf{T}\mathbf{t}$ is a $D \times 1$ vector then we simply compare terms and clearly

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{t} = \mathbf{X}^\mathsf{T}\mathbf{t} \tag{9}$$

The third term is a quadratic form in $\mathbf{w}$ and from section 2.3.2 of The Matrix Cookbook we have

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w} = \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \mathbf{X}^\mathsf{T}\mathbf{X}\right)\mathbf{w} = 2\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w} \tag{10}$$

This follows as $\mathbf{X}^\mathsf{T}\mathbf{X}$ is a $D \times D$ square-symmetric matrix and the transpose of a symmetric matrix is the matrix itself.

Collecting the terms from Equation (9) and (10) gives us the required expression for the vector of partial derivatives of MSE.

$$\frac{\partial MSE}{\partial \mathbf{w}} = -\frac{2}{N}\mathbf{X}^\mathsf{T}(\mathbf{t} - \mathbf{X}\mathbf{w}) \tag{11}$$

3 Matrix of Second Partial Derivatives:The Hessian

From the expression for the vector of partial derivatives of MSE we can see that the only term which is a function of $\mathbf{w}$ is $\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w}$ (neglecting the constant term $-\frac{2}{N}$.

From The Matrix Cookbook page 10, section 2.3.4 we can see that the $D \times D$ matrix of second partial derivatives of a quadratic form is simply $\mathbf{X}^\mathsf{T}\mathbf{X}$ as required.

4 Conclusion

It is good for the soul to convince oneself of the validity of the vector and matrix derivative we have employed from The Matrix Cookbook by working through a simple example longhand. Once is probably more than enough.