Easer Derivative

Thibaut Van Goethem

October 2021

1 Problem

Defined Lagrangian used for constrained optimization problem:

$$L = ||X - XB||_F^2 + \lambda \cdot ||B||_F^2 + 2 \cdot \gamma^T \cdot \mathrm{diag}(B) \tag{1}$$

As this is the lagrangian we know that we have the optimal solution when we set the derivative to 0

2 Frobenius norm

The Frobenius norm of a matrix is defined as the root of the sum of all squared elements of the matrix. this norm can also be written as:

$$||B||_F = tr(B^T B)^{\frac{1}{2}} \tag{2}$$

With tr being the trace operation of a matrix, this operation is the sum of all diagonal elements.
Note that in the used equations we always use the square of this norm which means that the power gets removed.

3 Derivative

$$0 = \frac{d}{dB}||X - XB||_F^2 + \lambda \cdot ||B||_F^2 + 2 \cdot \gamma^T \cdot \mathrm{diag}(B) \tag{3}$$

3.1 $\frac{d}{dB}||X - XB||_F^2$

$$\frac{d}{dB}||X - XB||_F^2 = \frac{d}{dB}\mathrm{tr}((X - XB)^T (X - XB)) \tag{4}$$

$$=> \frac{d}{dB}\text{tr}(X^TX - X^TXB - B^TX^TX + B^TX^TXB) \tag{5}$$

the trace operation is associative and commutative so we can split up this equation on the plus and minus operation, then we can solve each sub derivative separately.
See Matrix cookbook for derivatives of the trace operation of matrixes (section 2.5)

$$=> -X^TX - X^TX + X^TXB + X^TXB \tag{6}$$

$$=> -2(X^TX) + 2(X^TXB) \tag{7}$$

3.2 $\frac{d}{dB}2\lambda||B||_F^2$

See Matrix cookbook for derivatives of a single matrix frobenuis norm (section 2.7.1)

$$\frac{d}{dB}2\lambda||B||_F^2 = 2\lambda B \tag{8}$$

3.3 $\frac{d}{dB}2\gamma^T \cdot \textbf{diag}(B)$

For this derivative we note that the dotproduct between γ^T and diag(B) ends up being the sum of each time an element of gamma times a diagonal element of B. Then when we derive toward B then we need to calculate the partial towards each element of B. due to our dotproduct only containing diagonal elements of B it will be so that in the resulting matrix we only end with entries on the diagonal. And these entries will be the γ elemts the respective B elements was paired up with in the dotproduct.So we end up with a diagonal matrix of γ elements

$$\frac{d}{dB}2\gamma^T \cdot \text{diag}(B) = 2diagMatrix(\gamma) \tag{9}$$

3.4 Combining

$$0 = -2(X^TX) + 2(X^TXB) + 2\lambda B + 2diagMatrix(\gamma) \tag{10}$$

Note here is the step which explains why we need the 2 in the original lagrangian formula multiplying the lagrange multipliers.

$$0 = -X^TX + X^TXB + \lambda B + diagMatrix(\gamma) \tag{11}$$

$$0 = -X^TX + (X^TX + \lambda I)B + diagMatrix(\gamma) \tag{12}$$

$$-(X^T X + \lambda I)B = -X^T X + diagMatrix(\gamma) \tag{13}$$

$$(X^T X + \lambda I)B = X^T X - diagMatrix(\gamma) \tag{14}$$

$$B = (X^T X + \lambda I)^- 1 \cdot (X^T X - diagMatrix(\gamma)) \tag{15}$$

■