MAT-INF4130
Mandatory Assignment 2

Andreas Slyngstad

19. oktober 2016

We are presented the general problem
Find a k-dimensional subspace W of R m so that the sum of the (squared) distances from a set of n given points x1,x2,. . . ,xn in R m to W is as small as possible We define the *matrix of observations* **X** as

$$\mathbf{X} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}$$

1 Problem 1

First we assume that W is known such that $\{w_i\}_{i=1}^k$ is an orthonormal basis for W, and let $\{w_i\}_{i=k+1}^m$ be an orthonormal basis for $W^\perp$. We define the projection operator $proj_u v = \frac{\langle v,u \rangle}{\langle u,u \rangle}$ as the projection of a vector v onto the vector u.
Now using this operator to find the projection of the columns of the matrix of observations x_i on the space $W^\perp$ we get

$$\sum_{i=1}^{n} ||proj_{W^\perp} x_i||^2 = \sum_{i=1}^{n} \frac{\langle x_i^T, W^\perp \rangle}{\langle W^\perp, W^\perp \rangle} = \sum_{i=1}^{n} \sum_{i=k+1}^{m} \frac{\langle x_i^T, w_j \rangle}{\langle w_j, w_j \rangle}$$

Since $\{w_i\}_{i=k+1}^m$ spans an orthonormal basis for $W^\perp$, by definition $\langle w_i, w_i \rangle = 1$. From this we can rewrite the last sum in a more compact form as the frobenius norm of the matrix product $||X^T$ $\mathbf{W}||$, where $\mathbf{W}$ is the matrix with columns $\{w_i\}_{i=k+1}^m$

2 Problem 2

As introduced, if $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$ is and SVD of $\mathbf{X}$ then $\mathbf{X}^T = \mathbf{V}\Sigma^T\mathbf{U}^*$ we have the relation

$$||\mathbf{X}^T\mathbf{W}||_F^2 = ||\Sigma^T\mathbf{U}^*\mathbf{W}||_F^2 = \sum_{i=1}^{m}\sum_{j=k+1}^{m}\sigma_i^2\langle\mathbf{u}_i, \mathbf{w}_j\rangle^2 = \sum_{i=1}^{m}\sigma_i^2||proj_{w^\perp}\mathbf{u}_i||^2$$

From SVD we know that the unitary matrix $\mathbf{U}$ spans an orthonormal basis of $\mathbf{X}$ such that $(\mathbf{u}_1, \mathbf{u}_2....\mathbf{u}_k)$ for $k = rank(X)$. Now to minimize $||\mathbf{X}^T\mathbf{W}||_F^2$ we want in some sence to construct W such that W is in the same span as $\mathbf{X}$, hence we want to construct W such that W spans the same space as the orthonormal basis $\mathbf{U}$ of $\mathbf{X}$.

Since W is a k-dimensional subspace of $\mathbb{R}$, we can orientate W such that $\{w_i\}_{i=1}^k$ spans the same space as the components $\mathbf{u}_i$ for $i = 1, 2..k$, in other words $W = span(\mathbf{u}_1, ..., \mathbf{u}_k)$.
Then $\sum_{i=1}^k ||proj_{w^\perp}\mathbf{u}_i||^2 = 0$ (Mark the upper limit of the sum)
Now $\mathbf{u}_i$ for $i = k+1, k+2...m$ will lie in $span(W^\perp)$, and therefore $\sum_{i=k+1}^m ||proj_{w^\perp}\mathbf{u}_i||^2 = m - k$ since both $\{w_i\}_{i=k+1}^m$ and $\{u_i\}_{i=k+1}^m$ spans an orthonormal basis.

From this argumentation, finding the optimal subspace of W is equivalent to minimizing the system $\sum_{i=1}^m \sigma_i^2 x_i$ which will happen when the conditions $x_1 = \cdots = x_k = 0, \quad x_{k+1} = \cdots = x_m = 1$ and $\sum_{i=1}^m x_i = m - k$.

And finally since $||proj_{w^\perp}\mathbf{u}_i||^2 = 1$ for $i = k+1, k+2...m$,

$\sum_{i=k+1}^m \sigma_i^2||proj_{w^\perp}\mathbf{u}_i||^2 = \sum_{i=k+1}^m \sigma_i^2 = ||\mathbf{X}^T\mathbf{W}||_F^2$

3 Problem 3

In exercise 2 we introduced the a vector $\mathbf{c}$ to help find the best approximation for the matrix $\mathbf{X}$ on $\mathbf{W}$ such that the distance from $\mathbf{x}_i - \mathbf{c}$ to $\mathbf{W}$ is as small as possible. Using the results from exercise 2 we replace $\mathbf{x}_i$ with $\mathbf{x}_i - \mathbf{c}$. This relation is the same as replacing $\mathbf{X}$ with $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{c}(1\ \ 1\cdots\ \ 1)$ Using this relation we get

$$\begin{aligned}
||\tilde{\mathbf{X}}^T\mathbf{w}_i||^2 = (\tilde{\mathbf{X}}^T\mathbf{w}_i)^T(\tilde{\mathbf{X}}^T\mathbf{w}_i) = \mathbf{w}_i^T\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\mathbf{w}_i \\
\mathbf{w}_i^T(\mathbf{X} - \mathbf{c}(1\ \ 1\cdots 1))(\mathbf{X} - \mathbf{c}(1\ \ 1\cdots 1))^T\mathbf{w}_i \\
= \mathbf{w}_i^T(\mathbf{X} - \mathbf{c}(1\ \ 1\cdots 1))(\mathbf{X}^T - (1\ \ 1\cdots 1)^T\mathbf{c}^T)\mathbf{w}_i \\
= \mathbf{w}_i^T(\mathbf{X}\mathbf{X}^T - \mathbf{X}(1\ \ 1\cdots 1)^T\mathbf{c}^T - \mathbf{c}(1\ \ 1\cdots 1)\mathbf{X}^T + \mathbf{c}(1\ \ 1\cdots 1)(1\ \ 1\cdots 1)^T\mathbf{c}^T)\mathbf{w}_i
\end{aligned}$$

1. $\mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w}$ falls out directly from multiplication
2. From $\mathbf{w}_i^T\mathbf{c}(1\ \ 1\cdots 1)(1\ \ 1\cdots 1)^T\mathbf{c}^T\mathbf{w}_i$
   we observe that the matrix multiplication $\mathbf{c}(1\ \ 1\cdots 1)\mathbf{c}(1\ \ 1\cdots 1)^T$
   (which is a scalar $(1 \times n)(n \times 1)$) will just be sum n, and by rearranging the terms we get $n\mathbf{c}_i^T\mathbf{w}_i\mathbf{w}_i^T\mathbf{c}$
3. The two final terms can be rewritten into one term. By a closer look at the $\mathbf{w}^T\mathbf{c}(1\ \ 1\cdots 1)\mathbf{X}^T\mathbf{w}$ we change the order of multiplication and still get the equivalent expression
   $(\sum_{i=1}^n x_i^{(i)}\ \ \sum_{i=1}^n x_i^{(i)} \cdots\ \ \sum_{i=1}^n x_m^{(i)})\mathbf{w}_i\mathbf{w}_i^T\mathbf{c}$. Equivalent argument yields for the last term

And we find that

$$||\tilde{\mathbf{X}}^T\mathbf{w}_i||^2 = n\mathbf{c}_i^T\mathbf{w}_i\mathbf{w}_i^T\mathbf{c} - 2(\sum_{i=1}^n x_1^{(i)}\ \ \sum_{i=1}^n x_2^{(i)} \cdots\ \ \sum_{i=1}^n x_m^{(i)})\mathbf{w}_i\mathbf{w}_i^T\mathbf{c} + n\mathbf{c}_i^T\mathbf{w}_i\mathbf{w}_i^T\mathbf{c}$$

4 Problem 4

In general the Lagrange method or method of Lagrange multipliers we want to find a local maxima or minima of a function f, with some constraints evalued from a function g. We define a Lagrange function $\mathcal{L}$ such that

$$\mathcal{L}(x_1, x_2, ..., x_n, \lambda_1, ..., \lambda_n) = f(_1, x_2, ..., x_n) - \sum_{i=1}^{n} \lambda_i g_i(x_1, x_2, ..., x_n)$$

where λ_i yields some scalar value to each constraint. Here the funtion g is limited to the set of points such that $g(x_1, x_2, ...x_n) = 0$ Further from a mathematical evaluation these local maxima or minima occur when the function f and g are parallel, which is equivalent to that the gradient of f and g are paralell. Hence the problem is reduced on the form

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \nabla g_i(\mathbf{x}) \tag{1}$$

In our case these multiple constraints is defined as $\mathbf{w}_i^T \mathbf{c} = 0$ for $i = 1, .., k$ as set in exercise 3 by the assumtion that vector $\mathbf{c} \in W$. Further the function f, generalized to our problem is to minimize $||\tilde{\mathbf{X}}^T \mathbf{W}||_F^2$, and the gradient is taken with respect to $\mathbf{c}$, as this is the value of interest to find the best approximation.

Now by (1) we easily see that the left hand side is just the gradient of $||\tilde{\mathbf{X}}^T \mathbf{W}||_F^2$ with respect to $\mathbf{c}$ as defined in exercise 3. For the right hand side we get the result by

$$\frac{\partial}{\partial \mathbf{c}} g(\mathbf{x}) = \frac{\partial}{\partial \mathbf{c}} \mathbf{w}_i^T c = \mathbf{w}_i^T \quad \text{for i = 1, ..., k}$$

$$\sum_{i=1}^{k} \lambda_i \nabla g_i(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \mathbf{w}_i^T = (\mathbf{w}_1 \;\; \mathbf{w}_2 \cdots \mathbf{w}_k) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix}$$