Definitions from Statistics

In the following, $\vec{x}$ will represent a vector that holds N data values, and x_i are the elements of $\vec{x}$. Unless stated otherwise, assume that the norm is the 2−norm,

$$\|\vec{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

1. The (sample) mean, $\bar{x}$:

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j$$

2. The (sample) variance, s^2:

$$s^2 = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \bar{x})^2$$

   Notice that if $\bar{x} = 0$, then this could be written as

$$s^2 = \frac{1}{N-1}\|\vec{x}\|_2^2 = \frac{1}{N-1}\vec{x}^T\vec{x}$$

3. The (sample) standard deviation is the square root of the variance:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(x_j - \bar{x})^2}$$

   Notice that if $\bar{x} = 0$, then this could be written as

$$s = \frac{1}{N-1}\|\vec{x}\|$$

4. The sample covariance between data sets $\vec{x}$ and $\vec{y}$ is given below. We are comparing x_1 against y_1, etc., so in this case $\vec{x}$ and $\vec{y}$ must have the same number of elements.

$$\mathrm{Cov}(x, y) = s_{xy}^2 = \frac{1}{N-1}\sum_{k=1}^{N}(x_k - \bar{x})(y_k - \bar{y})$$

   NOTE: Using $N - 1$ gives what is called the "least biased estimator" of the covariance. Sometimes we want to use $1/N$ instead, and sometimes the scaling of the covariance will not matter- So we want to pay particular attention to the situation.

5. The correlation coefficient, ρ_{xy} is defined as:

$$\rho_{xy} = \frac{s_{xy}^2}{s_x s_y}$$

NOTE: If $\vec{x}$ and $\vec{y}$ have zero mean already, this formula could be written in linear algebra form:

$$\rho_{xy} = \frac{\vec{x}^T \vec{y}}{\|\vec{x}\| \, \|\vec{y}\|}$$

This shows that the correlation coefficient can be interpreted as $\cos(\theta)$.

6. **The covariance matrix.**

Instead of just two data sets, $\vec{x}$ and $\vec{y}$, we might have p of them. We can collect all of this data as a $p \times n$ matrix X, where we could think of X has p row vectors of data, each with n items or n column vectors of data, each with p items- As it turns out, it will not matter which way we think about it.

First, we mean subtract the matrix (double-centering works well!). Next, we compute the (scaled) covariance between the i^{th} column and the j^{th} column:

$$S_{ij}^2 = \frac{1}{p} \sum_{k=1}^{p} X_{k,i} X_{k,j}$$

Computing this over all i, j results in an $n \times n$ symmetrix matrix. It turns out that we can denote the covariance matrix easily as:

$$C = \frac{1}{p} X^T X$$

(where X has been mean-subtracted).

In Matlab:

```
[p,n]=size(X);
m=mean(X);
Xm=X-repmat(m,p,1);
C=(1/p)*X'*X;
```

Matlab also has the covariance function built-in (it will do the mean-subtraction as well): `C=cov(X)`