Lecture 22

Andrei Antonenko

March 16, 2005

1 Covariance and Correlation

If X and Y are random variable, their **correlation coefficient** is defined as:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbf{var}\,(X)\,\mathbf{var}\,(Y)}} \tag{1}$$

It can be proved that

$$-1 \le \rho(X,Y) \le 1. \tag{2}$$

Proof. Assume that variances of X and Y are σ_X^2 and σ_Y^2 respectively. We have

$$\begin{aligned}
0 &\le \mathbf{var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \\
&= \frac{\mathbf{var}\,(X)}{\sigma_X^2} + \frac{\mathbf{var}\,(Y)}{\sigma_Y^2} + \frac{2\,\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} \\
&= 1 + 1 + 2\rho(X,Y) \\
&= 2(1+\rho(X,Y)).
\end{aligned}$$

Therefore,

$$-1 \le \rho(X,Y).$$

Similarly,

$$\begin{aligned}
0 &\le \mathbf{var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \\
&= \frac{\mathbf{var}\,(X)}{\sigma_X^2} + \frac{\mathbf{var}\,(Y)}{\sigma_Y^2} - \frac{2\,\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} \\
&= 1 + 1 - 2\rho(X,Y) \\
&= 2(1-\rho(X,Y)).
\end{aligned}$$

Therefore,

$$\rho(X,Y) \le 1.$$

$\square$

If $\rho = 1$ ($\rho = -1$), that means that there exists a positive (negative) constant c such that

$$c(x - \mathbf{E}\,[X]) = (y - \mathbf{E}\,[Y]), \quad \text{for all pairs } (x,y). \tag{3}$$

Therefore, if $|\rho| = 1$, there is a precise linear dependency between X and Y.

Example 1.1. Consider n tosses of the coin, and let X and Y be the numbers of heads and tails respectively. Therefore, we have:

$$\begin{aligned} x + y &= n \\ \mathbf{E}[X] + \mathbf{E}[Y] &= n, \end{aligned}$$

and therefore,

$$x - \mathbf{E}[X] = -(y - \mathbf{E}[Y]).$$

Now,

$$\begin{aligned} \operatorname{Cov}(X, Y) &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}\left[-(X - \mathbf{E}[X])^2\right] \\ &= -\mathbf{var}(X). \end{aligned}$$

The correlation coefficient is

$$\begin{aligned} \rho(X, Y) &= \frac{\operatorname{Cov}(X, Y)}{\sqrt{\mathbf{var}(X)\,\mathbf{var}(Y)}} \\ &= \frac{-\mathbf{var}(X)}{\sqrt{\mathbf{var}(X)\,\mathbf{var}(X)}} \\ &= -1, \end{aligned}$$

since $\mathbf{var}(X) = \mathbf{var}(Y)$.

There is a formula for the variance of the sum of random variables:

$$\mathbf{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbf{var}(X_i) + 2 \sum_{i,j=1;i<j}^{n} \operatorname{Cov}(X_i, X_j). \tag{4}$$

Example 1.2. Consider n people who took off their hats, mixed them, and than took a random hat. What is the expectation of the number of people who got their own hat? What is the variance?

Let X_i be the random variable which is equal to 1 if an i-th person got his own hat and 0 otherwise. The total number of people who got their own hats is

$$X = X_1 + X_2 + \cdots + X_n.$$

Now, the probability that a person picks his own hat is $1/n$. Therefore, X_i is a Bernoulli random variable, taking the value 1 with probability $1/n$ and value 0 with probability $1 - 1/n$. Thus, $\mathbf{E}[X_i] = 1/n$. Now,

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n] = 1/n + 1/n + \cdots + 1/n = 1.$$

Since X_i is a Bernoulli random variable, $\mathbf{var}(X_i) = \frac{1}{n}\left(1 - \frac{1}{n}\right)$.

Now we will compute the variance of X. But first we need to know $\operatorname{Cov}(X_i, X_j)$. We have:

$$\begin{aligned} \operatorname{Cov}(X_i, X_j) &= \mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\,\mathbf{E}[X_j] \\ &= P(X_i = 1, X_j = 1) - P(X_i = 1)P(X_j = 1) \\ &= P(X_i = 1)P(X_j = 1 | X_i = 1) - P(X_i = 1)P(X_j = 1) \\ &= \frac{1}{n} \times \frac{1}{n-1} - \frac{1}{n} \times \frac{1}{n} \\ &= \frac{1}{n^2(n-1)}. \end{aligned}$$

Using the formula for the variance of the sum of random variables, we obtain:

$$\begin{aligned}
\mathbf{var}\,(X) &= \mathbf{var}\,(X_1 + X_2 + \cdots + X_n) \\
&= \sum_{i=1}^{n} \mathbf{var}\,(X_i) + 2 \sum_{i,j=1;i<j}^{n} \mathrm{Cov}(X_i, X_j) \\
&= n \times \frac{1}{n}\left(1 - \frac{1}{n}\right) + 2\frac{n(n-1)}{2} \times \frac{1}{n^2(n-1)} \\
&= 1.
\end{aligned}$$

2 Estimation

Assume X is a random variable. Assume we know the distribution of X, but we don't know the value the random variable X took. What would be the best guess? Intuitively, it should be equal to the expectation of the random variable. Now we will formalize this notion. We would like to find an estimate of X c, which minimizes the **expected squared error** $\mathbf{E}\,[(X-c)^2]$. Let $\mu = \mathbf{E}\,[X]$. We have:

$$\begin{aligned}
\mathbf{E}\left[(X-c)^2\right] &= \mathbf{E}\left[(X-\mu+\mu-c)^2\right] \\
&= \mathbf{E}\left[(X-\mu)^2\right] + 2\mathbf{E}\,[(X-\mu)(\mu-c)] + \mathbf{E}\left[(\mu-c)^2\right] \\
&= \mathbf{var}\,(X) + \mathbf{E}\left[(\mu-c)^2\right].
\end{aligned}$$

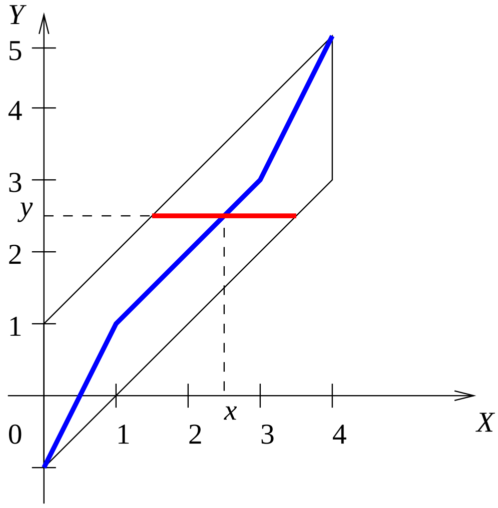
The first term in this expression does not depend on c, and the second term is minimized when $c = \mu$. Therefore, the mean squared error is minimized, when $c = \mu = \mathbf{E}\,[X]$.

Now assume we are given information about the value of the other random variable Y, more precisely, assume we know that $Y = y$. What would be the best guess for the value of X in this case? Since now we have a universe in which $Y = y$, by similar arguments as in the previous paragraph, we can see that in this case our best guess will be $\mathbf{E}\,[X|Y=y]$. Formally, we can state it as follows. The value of c which minimizes the expression $\mathbf{E}\,[(c-X)^2|Y=y]$ is $c = \mathbf{E}\,[X|Y=y]$.

Now let's consider an example.

Example 2.1. Assume that we know that X is a random variable, uniformly distributed from 0 to 4. Imagine the value of X is transmitted over the phone line, and the actual value received is corrupted by noise W, which is uniformly distributed from -1 to 1. Therefore, instead of actual value of X, we receive $Y = X + W$. Assume, the received value is $Y = y$. What would be the best guess for the value of X in this case?

Since Y is equal to $X + W$, we can see that Y is uniformly distributed on $[x-1, x+1]$. The corresponding picture is given below.



Now, for each value of $Y = y$ we can see the interval, where X is distributed (marked by a red line on the graph). Since X is uniform on this interval, we should take a center point of it, to get the optimal prediction of X, which is $\mathbf{E}[X|Y = y]$. The blue line on the graph shows the $\mathbf{E}[X|Y = y]$ for all possible values of $Y = y$. Therefore, knowing $Y = y$, we can find the value $X = x$, which will be our best guess in predicting the real value.