Machine Learning for Engineering and Science Applications

Variance

Covariance

Variance

- Random variables, by definition, result in different outcomes
- The variation in random variables is captured by their distribution
- Expectation gives mean/average/expected value of the random variable given the distribution
- **Variance** gives the *variation from the expected value*
- Variance also measures amount of fluctuation of the variable

  Examples:
  - Variance in returns on a certain investment in the market (Risk measure)
  - Variance in rainfall during the coming monsoon

Univariate variance

- The **variance** and its square root, **standard deviation** of some function $f(x)$ with respect to a probability distribution $P(x)$ measure how much the value of $f(x)$ varies for various samples when x is drawn from P

Denoted by $\mathbb{V}_{x\sim P}[f(x)]$

- If P is clear from the context $\mathbb{V}_x[f(x)]$
- If x is also clear from the context $\mathbb{V}[f(x)]$
- **Usually**, simply denoted as $\boxed{\mathbb{V}[f] \text{ or } \mathrm{Var}[f]}$

- Mathematically,

$$\boxed{\mathbb{V}[f(x)] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]}$$
$$= \mathbb{E}[\left(f(x) - \overline{f(x)}\right)^2]$$

The standard deviation is given as $\sigma[x] = \sqrt{\mathbb{V}[x]}$

Covariance (univariate)

- Note that $\mathbb{V}[x] = \mathbb{E}[(x - \overline{x})^2] = \mathbb{E}[(x - \overline{x})(x - \overline{x})]$

- This notion can be generalized to a pair of variables x and y to find the **covariance**

$$\mathrm{Cov}[\mathrm{x}, \mathrm{y}] = \mathbb{E}[(x - \overline{x})(y - \overline{y})] = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])]$$

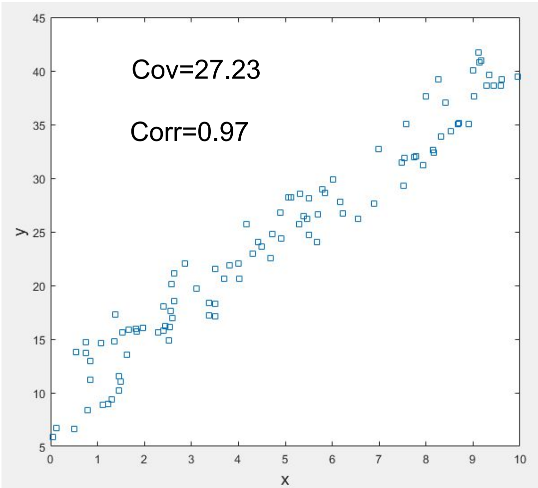
- Similarly, $\boxed{\mathrm{Cov}[f(\mathrm{x}), g(\mathrm{y})] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])(g(y) - \mathbb{E}[g(y)]\,)]}$

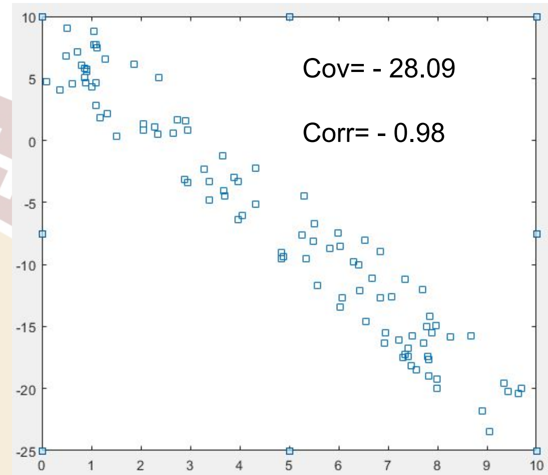
- A related quantity is the correlation, defined as

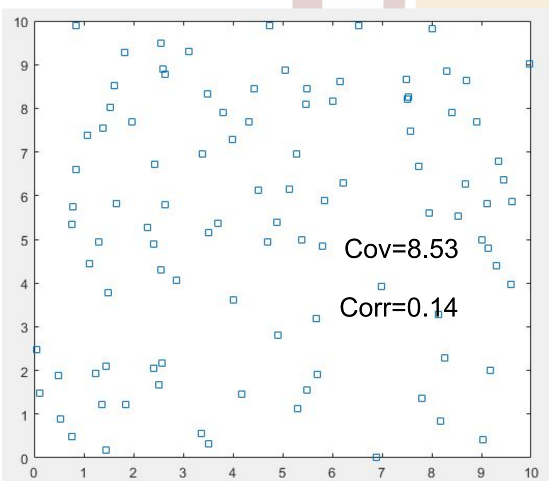
$$\boxed{\mathrm{corr}[x, y] = \frac{\mathrm{Cov}[x,y]}{\sqrt{\mathbb{V}[x]\mathbb{V}[y]}}}$$

- Measures how linearly correlated the two random variables are
- Note $\mathrm{Cov}[x, x] = \mathrm{Var}[x]$ and $\mathrm{corr}[x, x] = 1$

Interpreting Covariance and correlation







- Positive covariance means when x increases, y is expected to increase too.
- Negative covariance means when x increases, y is expected to decrease

- Correlation close to 1 means strongly, positively correlated
- Correlation close to -1 means strongly, negatively correlated

- Correlation close to 0 means no (linear) correlation

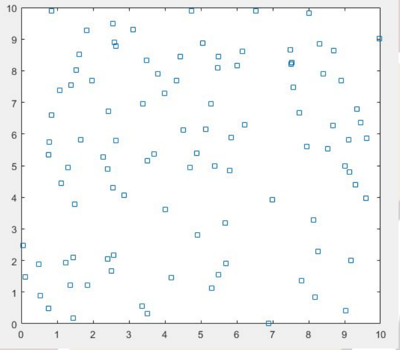
Covariance simplification

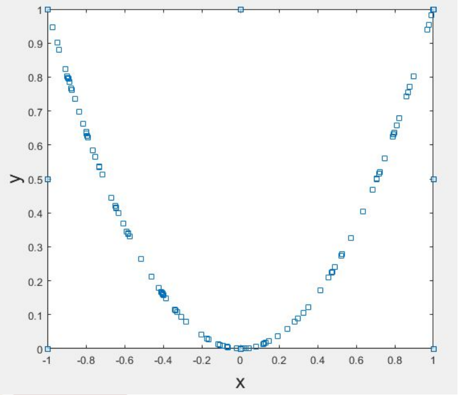
- $$\boxed{\mathrm{Cov}[\mathrm{x},\mathrm{y}] = \mathbb{E}[(x-\overline{x})(y-\overline{y})] = \mathbb{E}[(x-\mathbb{E}[x])(y-\mathbb{E}[y])]}$$
- This can be simplified as $\mathrm{Cov}[\mathrm{x},\mathrm{y}] = \overline{xy} - \overline{x}\,\overline{y} = \boxed{\mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]}$
- Proof : We use the fact that $\overline{\overline{x}} = \overline{x}$ i.e. $\mathbb{E}\big[\mathbb{E}[x]\big] = \mathbb{E}[x]$

For $x = y$, this relation gives $\boxed{\mathrm{Var}[x] = \mathbb{E}[x^2] - (\mathbb{E}[x])^2}$

Independence and covariance

- When x, y are two independent random variables, $\mathrm{Cov}[x, y] = 0$





- However, the converse is not true
  - That is, $\mathrm{Cov}[x, y] = 0$ **need not mean** x, y are independent
- Example – Let x be a random variable with $\mathbb{E}[x] = 0$ and $\mathbb{E}[x^3] = 0$
- Let $y = x^2$ be another random variable
- Then, $\mathrm{Cov}[\mathrm{x,y}] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] = \mathbb{E}[x^3] - \mathbb{E}[x]\mathbb{E}[x^2] = 0$
  - That is, covariance is zero even though the variables are not independent
  - It turns out zero covariance only means that there is no *linear* relationship

Covariance matrix

- If $\boldsymbol{x} \in \mathbb{R}^n$ is a vector, then it is often useful to know pairwise covariances between all pairs of components
  - Think of the *x* being the components of an input vector such as pixels of an image

- That is, define $\mathrm{Cov}[\boldsymbol{x}, \boldsymbol{x}]_{i,j} = \mathrm{Cov}[\boldsymbol{x}_i, \boldsymbol{x}_j]$

$$\mathrm{Cov}[x, x] = \begin{bmatrix} \mathrm{Cov}[x_1, x_1] & \mathrm{Cov}[x_1, x_2] & \cdots & \mathrm{Cov}[x_1, x_n] \\ \mathrm{Cov}[x_2, x_1] & \mathrm{Cov}[x_2, x_2] & \dots & \mathrm{Cov}[x_2, x_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[x_n, x_1] & \mathrm{Cov}[x_n, x_2] & \cdots & \mathrm{Cov}[x_n, x_n] \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- Note that the diagonal elements are simply the variances of individual components, that is, $\mathrm{Cov}[x_i, x_i] = \mathrm{Var}[x_i]$