Aalto University
School of Electrical Engineering

CS-C3240 – Machine Learning D

Feature Engineering

Stephan Sigg

Department of Communications and Networking
Aalto University, School of Electrical Engineering
stephan.sigg@aalto.fi

Version 1.0, February 14, 2022

Variance and Covariance

For two features $\mathcal{X}_1, \mathcal{X}_2$, and the arithmetic mean $E[\cdot]$, consider sets of measurements with zero mean:

$$\overrightarrow{x_1} = \{x_1^{(1)}, \dots, x_1^{(n)}\}$$
$$\overrightarrow{x_2} = \{x_2^{(1)}, \dots, x_2^{(n)}\}$$

Variance and Covariance

For two features $\mathcal{X}_1, \mathcal{X}_2$, and the arithmetic mean $E[\cdot]$, consider sets of measurements with zero mean:

$$\overrightarrow{x_1} = \{x_1^{(1)}, \ldots, x_1^{(n)}\}$$
$$\overrightarrow{x_2} = \{x_2^{(1)}, \ldots, x_2^{(n)}\}$$

Variance: $E[(\overrightarrow{x_j} - E[\overrightarrow{x_j}])^2] = E[(\overrightarrow{x_j} - \mu)^2]$

Variance and Covariance

For two features $\mathcal{X}_1, \mathcal{X}_2$, and the arithmetic mean $E[\cdot]$, consider sets of measurements with zero mean:

$$\overrightarrow{x_1} = \{x_1^{(1)}, \dots, x_1^{(n)}\}$$
$$\overrightarrow{x_2} = \{x_2^{(1)}, \dots, x_2^{(n)}\}$$

Variance: $E[(\overrightarrow{x_j} - E[\overrightarrow{x_j}])^2] = E[(\overrightarrow{x_j} - \mu)^2]$

$\xrightarrow{\text{zero mean}} E[\overrightarrow{x_j} \cdot \overrightarrow{x_j}]$

Variance and Covariance

For two features $\mathcal{X}_1, \mathcal{X}_2$, and the arithmetic mean $E[\cdot]$, consider sets of measurements with zero mean:

$$\overrightarrow{x_1} = \{x_1^{(1)}, \ldots, x_1^{(n)}\}$$
$$\overrightarrow{x_2} = \{x_2^{(1)}, \ldots, x_2^{(n)}\}$$

Variance: $E[(\overrightarrow{x_j} - E[\overrightarrow{x_j}])^2] = E[(\overrightarrow{x_j} - \mu)^2]$

$\xrightarrow{\text{zero mean}} E[\overrightarrow{x_j} \cdot \overrightarrow{x_j}]$

$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} (\overrightarrow{x_j} \cdot \overrightarrow{x_j})$

Variance and Covariance

For two features $\mathcal{X}_1, \mathcal{X}_2$, and the arithmetic mean $E[\cdot]$, consider sets of measurements with zero mean:

$$\overrightarrow{x_1} = \{x_1^{(1)}, \ldots, x_1^{(n)}\}$$
$$\overrightarrow{x_2} = \{x_2^{(1)}, \ldots, x_2^{(n)}\}$$

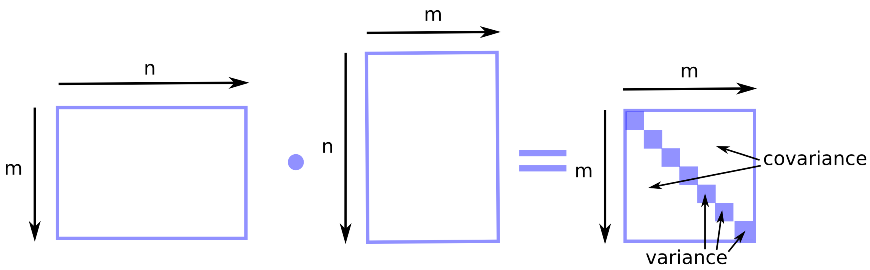
Variance: $E[(\overrightarrow{x_j} - E[\overrightarrow{x_j}])^2] = E[(\overrightarrow{x_j} - \mu)^2]$

$\xrightarrow{\text{zero mean}} E[\overrightarrow{x_j} \cdot \overrightarrow{x_j}]$

$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} (\overrightarrow{x_j} \cdot \overrightarrow{x_j})$

Covariance: $E[(\overrightarrow{x_i} - E[\overrightarrow{x_i}])(\overrightarrow{x_j} - E[\overrightarrow{x_j}])]$

Covariance matrix



Questions?

Stephan Sigg
stephan.sigg@aalto.fi

Si Zuo
si.zuo@aalto.fi

Literature

- C.M. Bishop: Pattern recognition and machine learning, Springer, 2007.
- R.O. Duda, P.E. Hart, D.G. Stork: Pattern Classification, Wiley, 2001.