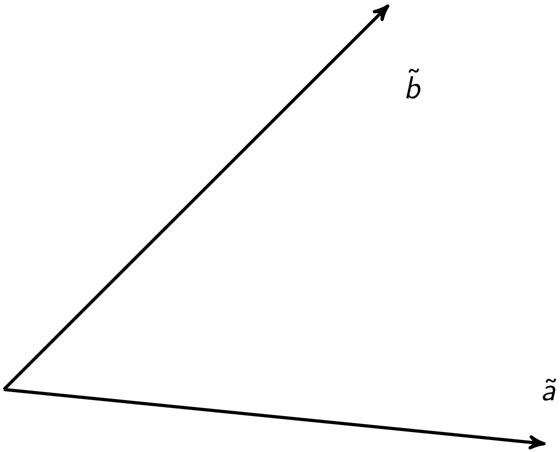
A Geometric Analysis Of Covariance

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Goal: Interpret random variables as vectors



Vectors

Objects that admit two operations:

1. Vector sum that is commutative and associative

$$v_1 + v_2 = v_2 + v_1$$
$$(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$$

2. Multiplication between scalars and vectors that is associative

$$\alpha_1(\alpha_2 v) = (\alpha_1 \alpha_2) v$$

Vector space

A set of vectors $\mathcal{V}$ is a valid vector space if:

- For any $v \in \mathcal{V}$ and any scalar $\beta \in \mathbb{R}$, $\beta v \in \mathcal{V}$
- For any $v_1, v_2 \in \mathcal{V}$, $v_1 + v_2 \in \mathcal{V}$
- There exists a zero vector 0 such that $v + 0 = v$ for any $v \in \mathcal{V}$
- For any $v \in \mathcal{V}$ there exists an additive inverse $-v$ such that $v + (-v) = 0$

Vector space of random variables

We consider $\tilde{a} = \tilde{b}$ if

$$\mathrm{P}\left(\tilde{a} = \tilde{b}\right) = 1$$

Let $\mathcal{R}$ be the set of random variables associated to a probability space

- For any $\tilde{a} \in \mathcal{R}$ and any scalar $\beta \in \mathbb{R}$, $\beta\tilde{a} \in \mathcal{R}$
- For any $\tilde{a}, \tilde{b} \in \mathcal{R}$, $\tilde{a} + \tilde{b} \in \mathcal{R}$
- The random variable $\tilde{0}$ that is equal to zero with probability one satisfies $\tilde{a} + \tilde{0} = \tilde{a}$ for any $\tilde{a} \in \mathcal{R}$
- For any $\tilde{a} \in \mathcal{R}$, $-\tilde{a} \in \mathcal{R}$ satisfies $\tilde{a} + (-\tilde{a}) = \tilde{0}$

Inner product

An inner product $\langle \cdot, \cdot \rangle$ on a vector space $\mathcal{V}$ is

- Symmetric: for any $v_1, v_2 \in \mathcal{V}$,

$$\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$$

- Linear: for any $\beta \in \mathbb{R}$ and any $v_1, v_2, v_3 \in \mathcal{V}$

$$\langle \beta\, v_1, v_2 \rangle = \beta \langle v_2, v_1 \rangle$$
$$\langle v_1 + v_2, v_3 \rangle = \langle v_1, v_3 \rangle + \langle v_2, v_3 \rangle$$

- Positive semidefinite: $\langle v, v \rangle$ is nonnegative for all $v \in \mathcal{V}$ and if $\langle v, v \rangle = 0$ then $v = 0$

Covariance as an inner product

The covariance between zero-mean random variables is

- Symmetric:

$$\begin{aligned}\mathrm{Cov}[\tilde{a}, \tilde{b}] &:= \mathrm{E}\left[(\tilde{a} - \mathrm{E}[\tilde{a}])(\tilde{b} - \mathrm{E}[\tilde{b}])\right] \\ &= \mathrm{E}\left[(\tilde{b} - \mathrm{E}[\tilde{b}])(\tilde{a} - \mathrm{E}[\tilde{a}])\right] = \mathrm{Cov}[\tilde{b}, \tilde{a}]\end{aligned}$$

- Linear:

$$\begin{aligned}\mathrm{Cov}[\beta\tilde{a}, \tilde{b}] &:= \mathrm{E}\left[(\beta\tilde{a} - \mathrm{E}[\beta\tilde{a}])(\tilde{b} - \mathrm{E}[\tilde{b}])\right] \\ &= \beta\mathrm{E}\left[(\tilde{b} - \mathrm{E}[\tilde{b}])(\tilde{a} - \mathrm{E}[\tilde{a}])\right] = \beta\mathrm{Cov}[\tilde{b}, \tilde{a}] \\ \mathrm{Cov}[\tilde{a}_1 + \tilde{a}_2, \tilde{b}] &:= \mathrm{E}[(\tilde{a}_1 + \tilde{a}_2)\tilde{b}] - \mathrm{E}[\tilde{a}_1 + \tilde{a}_2]\mathrm{E}[\tilde{b}] \\ &= \mathrm{E}[\tilde{a}_1\tilde{b}] - \mathrm{E}[\tilde{a}_1]\mathrm{E}[\tilde{b}] + \mathrm{E}[\tilde{a}_2\tilde{b}] - \mathrm{E}[\tilde{a}_2]\mathrm{E}[\tilde{b}] \\ &= \mathrm{Cov}[\tilde{a}_1, \tilde{b}] + \mathrm{Cov}[\tilde{a}_2, \tilde{b}]\end{aligned}$$

- Positive semidefinite: $\mathrm{E}[\tilde{a}^2] = 0$ implies $\mathrm{P}(\tilde{a} = 0) = 1$

Norm of a vector

The norm is the *length* of the vector

$$||v|| := \sqrt{\langle v, v \rangle}$$

For a zero-mean random variable

$$\begin{aligned} ||\tilde{a}|| &:= \sqrt{\mathrm{Cov}[\tilde{a}, \tilde{a}]} \\ &= \sqrt{\mathrm{Var}[\tilde{a}]} = \sigma_{\tilde{a}} \end{aligned}$$

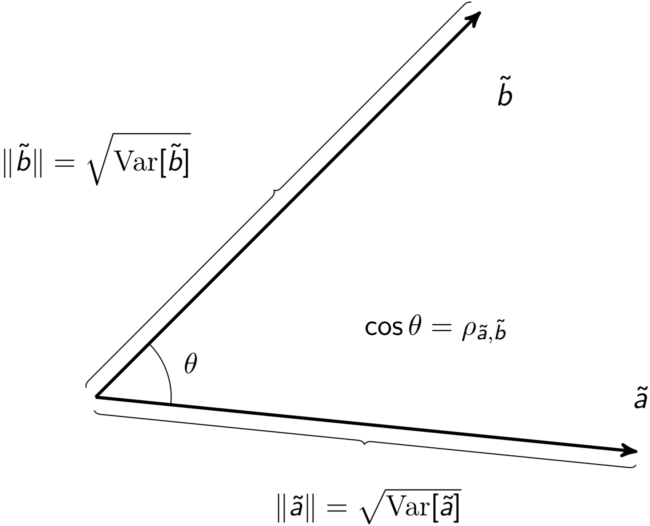
Angle between vectors

The cosine of the angle between two vectors is equal to the normalized inner product

For zero-mean random variables

$$\begin{aligned}
\cos\theta &= \frac{\langle \tilde{a}, \tilde{b} \rangle}{||\tilde{a}||\,\|\tilde{b}\|} \\
&= \frac{\mathrm{Cov}[\tilde{a}, \tilde{b}]}{\sqrt{\mathrm{Var}[\tilde{a}]\mathrm{Var}[\tilde{b}]}} \\
&= \rho_{\tilde{a},\tilde{b}}
\end{aligned}$$

Random variables as vectors



$-1 \leq \cos\theta \leq 1$ $\qquad$ $-1 \leq \rho_{\tilde{a},\tilde{b}} \leq 1$

If $\cos\theta > 0$ vectors point in the same direction

If $\rho_{\tilde{a},\tilde{b}} > 0$ $\tilde{a}$ and $\tilde{b}$ are positively correlated

If $\cos\theta < 0$ vectors point in opposite directions

If $\rho_{\tilde{a},\tilde{b}} < 0$ $\tilde{a}$ and $\tilde{b}$ are negatively correlated

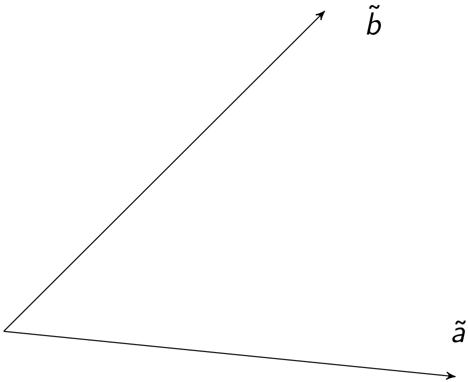
If $\cos\theta = \pm 1$ vectors are collinear

If $\rho_{\tilde{a},\tilde{b}} = \pm 1$ $\tilde{a}$ and $\tilde{b}$ are completely linearly dependent

If $\cos\theta = 0$ vectors are orthogonal

If $\rho_{\tilde{a},\tilde{b}} = 0$ $\tilde{a}$ and $\tilde{b}$ are uncorrelated

Simple linear regression

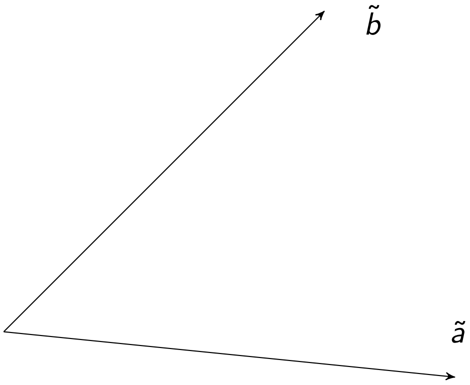


Simple linear regression

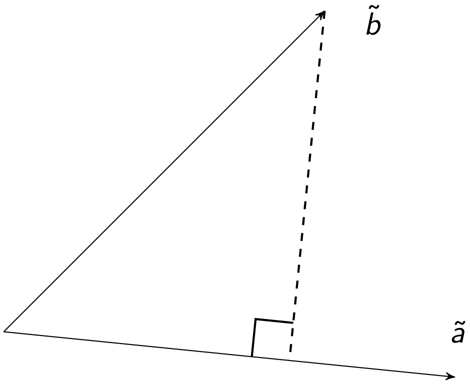
Mean squared error is squared distance

$$\begin{aligned}\mathrm{E}\left[(\tilde{b} - \beta \tilde{a})^2\right] &= \mathrm{Var}\left[\tilde{b} - \beta \tilde{a}\right] \\ &= \left|\left|\tilde{b} - \beta \tilde{a}\right|\right|^2\end{aligned}$$

Vector collinear with $\tilde{a}$ closest to $\tilde{b}$?



Orthogonal projection



Orthogonal projection

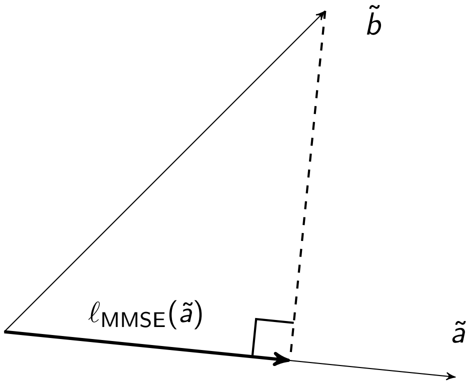
$$\langle \tilde{a}, \tilde{b} - \beta \tilde{a} \rangle = 0$$

$$\beta \, ||\tilde{a}||^2 = \langle \tilde{a}, \tilde{b} \rangle$$

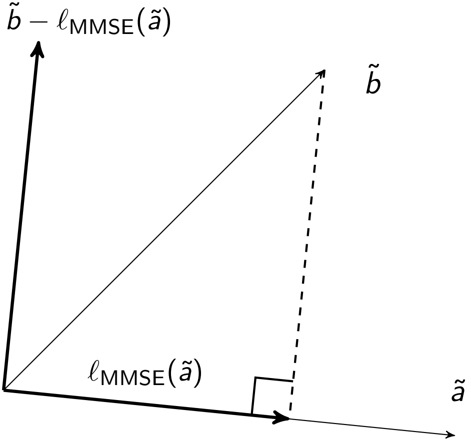
$$\begin{aligned} \beta &= \frac{\langle \tilde{a}, \tilde{b} \rangle}{||\tilde{a}||^2} \\ &= \frac{\mathrm{Cov}\,[\tilde{a}, \tilde{b}]}{\mathrm{Var}\,[\tilde{a}]} \\ &= \rho_{\tilde{a},\tilde{b}} \sqrt{\frac{\mathrm{Var}[\tilde{b}]}{\mathrm{Var}[\tilde{a}]}} \end{aligned}$$

$$\beta \tilde{a} = \ell_{\mathrm{MMSE}}(\tilde{a})$$

Linear minimum MSE estimator



Residual



$\tilde{b} - \ell_{\text{MMSE}}(\tilde{a})$
$\tilde{b}$
$\ell_{\text{MMSE}}(\tilde{a})$
$\tilde{a}$

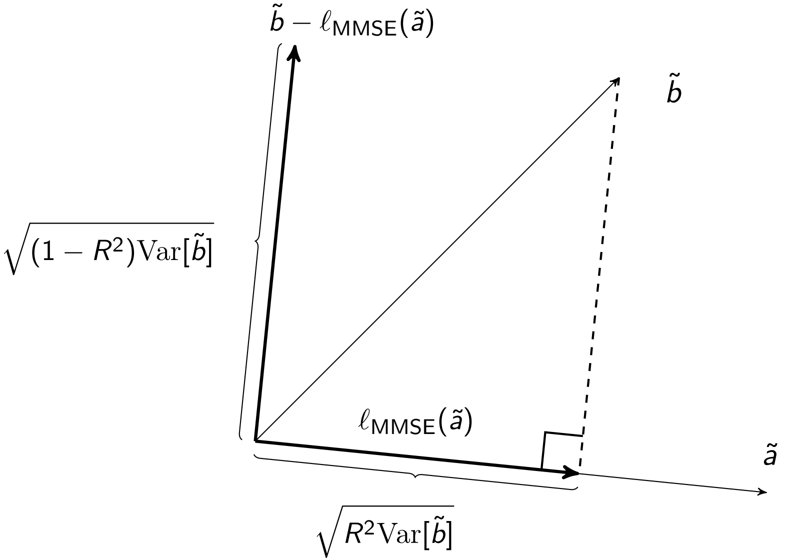
Pythagoras' theorem

$$\begin{aligned}
\mathrm{Var}[\tilde{b}] = ||\tilde{b}||^2 &= ||\ell_{\mathsf{MMSE}}(\tilde{a})||^2 + ||\tilde{b} - \ell_{\mathsf{MMSE}}(\tilde{a})||^2 \\
&= \mathrm{Var}\left[\ell_{\mathsf{MMSE}}(\tilde{a})\right] + \mathrm{Var}\left[\tilde{b} - \ell_{\mathsf{MMSE}}(\tilde{a})\right]
\end{aligned}$$

Length of projection

$$\begin{aligned}
||\ell_{\mathsf{MMSE}}(\tilde{a})||^2 &= ||\beta \tilde{a}||^2 \\
&= \beta^2 \, ||\tilde{a}||^2 \\
&= \frac{\rho_{\tilde{a},\tilde{b}}^2 \mathrm{Var}[\tilde{b}]^2}{\mathrm{Var}[\tilde{a}]^2} \mathrm{Var}[\tilde{a}]^2 \\
&= \rho_{\tilde{a},\tilde{b}}^2 \mathrm{Var}[\tilde{b}]^2
\end{aligned}$$

Decomposition of variance



What have we learned

Interpretation of:

- Random variables as vectors
- Covariance as inner product
- Correlation coefficient as cosine of angle
- Simple linear regression as orthogonal projection