Covariance matrix

DS GA 1002 Probability and Statistics for Data Science

Carlos Fernandez-Granda

Prerequisites

Linear algebra (vectors, inner products, projections)

Multivariate random variables

Expectation

Goals

Learn how to describe the average behavior of multivariate data using the covariance matrix

Mean of a random vector

Data x modeled as random vector $\tilde{x}$

The d-dimensional mean of a random vector $\tilde{x}$ is

$$\mathrm{E}(\tilde{x}) := \begin{bmatrix} \mathrm{E}\left(\tilde{x}[1]\right) \\ \mathrm{E}\left(\tilde{x}[2]\right) \\ \cdots \\ \mathrm{E}\left(\tilde{x}[d]\right) \end{bmatrix}$$

Mean of a random matrix

The mean of a $d_1 \times d_2$ matrix with random entries $\tilde{X}$ is

$$\mathrm{E}(\tilde{X}) := \begin{bmatrix} \mathrm{E}\left(\tilde{X}[1,1]\right) & \mathrm{E}\left(\tilde{X}[1,2]\right) & \cdots & \mathrm{E}\left(\tilde{X}[1,d_2]\right) \\ \mathrm{E}\left(\tilde{X}[2,1]\right) & \mathrm{E}\left(\tilde{X}[2,2]\right) & \cdots & \mathrm{E}\left(\tilde{X}[2,d_2]\right) \\ & & \cdots & \\ \mathrm{E}\left(\tilde{X}[d_1,1]\right) & \mathrm{E}\left(\tilde{X}[d_1,2]\right) & \cdots & \mathrm{E}\left(\tilde{X}[d_1,d_2]\right) \end{bmatrix}$$

Linearity of expectation for random vectors and matrices

For any d_1-dimensional random vector $\tilde{x}$, $d_1 \times d_2$ random matrix $\tilde{X}$, $A \in \mathbb{R}^{m \times d_1}$, $b \in \mathbb{R}^m$ and $B \in \mathbb{R}^{m \times d_2}$

$$\mathrm{E}(A\tilde{x} + b) = A\mathrm{E}(\tilde{x}) + b$$

$$\mathrm{E}(A\tilde{X} + B) = A\mathrm{E}(\tilde{X}) + B$$

Proof:

$$\begin{aligned}
\mathrm{E}(A\tilde{x} + b)[i] &= \mathrm{E}\left((A\tilde{x} + b)[i]\right) \\
&= \mathrm{E}\left(\sum_{j=1}^{d} A[i,j]\tilde{x}[j] + b[i]\right) \\
&= \sum_{j=1}^{d} A[i,j]\mathrm{E}\left(\tilde{x}[j]\right) + b[i] \\
&= (A\mathrm{E}(\tilde{x}) + b)[i]
\end{aligned}$$

Sample mean of multivariate data

The sample mean of a d-dimensional dataset $X := \{x_1, x_2, \ldots, x_n\}$ is

$$\mu_X := \frac{\sum_{i=1}^{n} x_i}{n}$$

Variance of linear combination $v^T \tilde{x}$?

For any vector v, the variance of $v^T \tilde{x}$ is given by

$$\begin{aligned}
\mathrm{Var}\left(v^T \tilde{x}\right) &= \mathrm{E}\left((v^T \tilde{x} - \mathrm{E}(v^T \tilde{x}))^2\right) \\
&= \mathrm{E}\left((v^T c(\tilde{x}))^2\right) \\
&= \mathrm{E}\left(v^T c(\tilde{x}) c(\tilde{x})^T v\right) \\
&= v^T \mathrm{E}\left(c(\tilde{x}) c(\tilde{x})^T\right) v
\end{aligned}$$

where $c(\tilde{x}) := \tilde{x} - \mathrm{E}(\tilde{x})$

Covariance matrix

The covariance matrix of a random vector $\tilde{x}$ is defined as

$$\Sigma_{\tilde{x}} := \mathrm{E}\left(c(\tilde{x})c(\tilde{x})^T\right)$$

$$= \begin{bmatrix} \mathrm{Var}\left(\tilde{x}[1]\right) & \mathrm{Cov}\left(\tilde{x}[1], \tilde{x}[2]\right) & \cdots & \mathrm{Cov}\left(\tilde{x}[1], \tilde{x}[d]\right) \\ \mathrm{Cov}\left(\tilde{x}[1], \tilde{x}[2]\right) & \mathrm{Var}\left(\tilde{x}[2]\right) & \cdots & \mathrm{Cov}\left(\tilde{x}[2], \tilde{x}[d]\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(\tilde{x}[1], \tilde{x}[d]\right) & \mathrm{Cov}\left(\tilde{x}[2], \tilde{x}[d]\right) & \cdots & \mathrm{Var}\left(\tilde{x}[d]\right) \end{bmatrix}$$

Variance of linear combination $v^T \tilde{x}$

$$\begin{aligned}
\mathrm{Var}\left(v^T \tilde{x}\right) &= \mathrm{E}\left((v^T \tilde{x} - \mathrm{E}(v^T \tilde{x}))^2\right) \\
&= \mathrm{E}\left((v^T c(\tilde{x}))^2\right) \\
&= v^T \mathrm{E}\left(c(\tilde{x}) c(\tilde{x})^T\right) v \\
&= \mathrm{E}\left(v^T c(\tilde{x}) c(\tilde{x})^T v\right) \\
&= v^T \Sigma_{\tilde{x}} v
\end{aligned}$$

Cheese sandwich

Ingredients: bread, local cheese, and imported cheese

Model for price of ingredients: random vector $\tilde{x}$ (cents/gram) with covariance matrix

$$\Sigma_{\tilde{x}} = \begin{bmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0 \\ 0 & 0 & 1.2 \end{bmatrix}$$

Two recipes:

1. 100g of bread, 50g of local cheese, and 50g of imported cheese
2. 100g of bread, 100g of local cheese, and no imported cheese

Which has higher standard deviation?

Recipe 1

$$\begin{aligned}
\sigma_{100\tilde{x}[1]+50\tilde{x}[2]+50\tilde{x}[3]} &= \sqrt{\begin{bmatrix} 100 & 50 & 50 \end{bmatrix} \Sigma_{\tilde{x}} \begin{bmatrix} 100 \\ 50 \\ 50 \end{bmatrix}} \\
&= \sqrt{\begin{bmatrix} 100 & 50 & 50 \end{bmatrix} \begin{bmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0 \\ 0 & 0 & 1.2 \end{bmatrix} \begin{bmatrix} 100 \\ 50 \\ 50 \end{bmatrix}} \\
&= 153 \text{ cents}
\end{aligned}$$

Recipe 2

$$\begin{aligned}
\sigma_{100\tilde{x}[1]+100\tilde{x}[2]} &= \sqrt{\begin{bmatrix}100 & 100 & 0\end{bmatrix} \Sigma_{\tilde{x}} \begin{bmatrix}100\\100\\0\end{bmatrix}} \\
&= \sqrt{\begin{bmatrix}100 & 100 & 0\end{bmatrix} \begin{bmatrix}1 & 0.8 & 0\\0.8 & 1 & 0\\0 & 0 & 1.2\end{bmatrix} \begin{bmatrix}100\\100\\0\end{bmatrix}} \\
&= 164 \text{ cents}
\end{aligned}$$

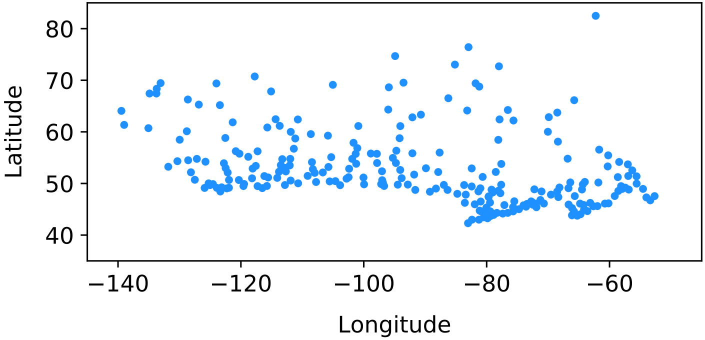
Sample covariance matrix

The sample covariance matrix of $X := \{x_1, x_2, \ldots, x_n\}$ is

$$\Sigma_X := \frac{1}{n}\sum_{i=1}^{n} c(x_i)c(x_i)^T$$
$$= \begin{bmatrix} \sigma^2_{X[1]} & \sigma_{X[1],X[2]} & \cdots & \sigma_{X[1],X[d]} \\ \sigma_{X[1],X[2]} & \sigma^2_{X[2]} & \cdots & \sigma_{X[2],X[d]} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X[1],X[d]} & \sigma_{X[2],X[d]} & \cdots & \sigma^2_{X[d]} \end{bmatrix}$$

$c(x_i) := x_i - \mu_X$
$X[j] := \{x_1[j], \ldots, x_n[j]\}$
$\sigma^2_{X[i]}$ is the sample variance of $X[i]$
$\sigma_{X[i],X[j]}$ is the sample covariance of $X[i]$ and $X[j]$

Cities in Canada



Sample covariance matrix:

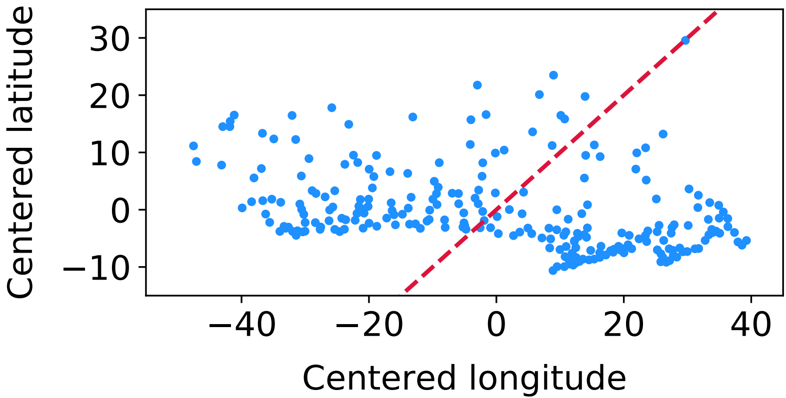
$$\Sigma_X = \begin{bmatrix} 524.9 & -59.8 \\ -59.8 & 53.7 \end{bmatrix}$$

Sample variance of linear combinations

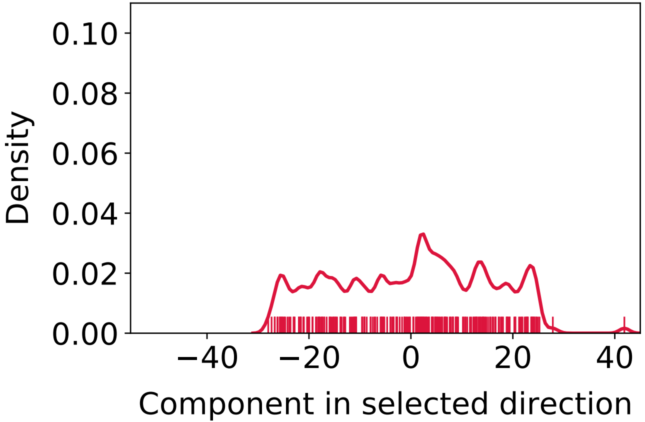
For $X = \{x_1, \ldots, x_n\}$ of d-dimensional data and $v \in \mathbb{R}^d$, let

$$
\begin{aligned}
X_v &:= \{\langle v, x_1 \rangle, \ldots, \langle v, x_n \rangle\} \\
\sigma^2_{X_v} &:= \frac{1}{n} \sum_{i=1}^{n} \left(v^T x_i - \mu_{X_v} \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(v^T x_i - \frac{1}{n} \sum_{j=1}^{n} v^T x_j \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(v^T \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right) \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} (v^T c(x_i))^2 \\
&= v^T \left(\frac{1}{n} \sum_{i=1}^{n} c(x_i) c(x_i)^T \right) v = v^T \Sigma_X v
\end{aligned}
$$

Projection onto a fixed direction



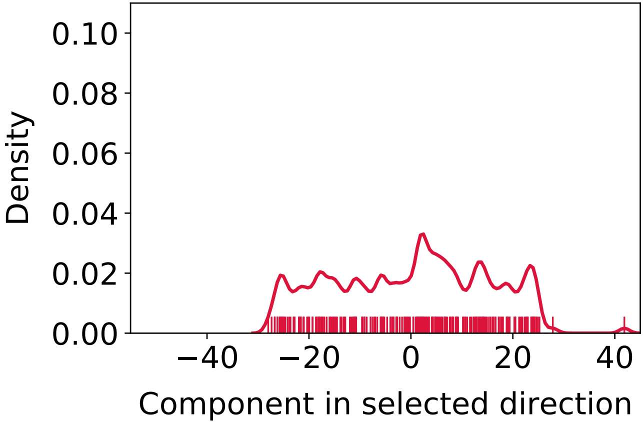
Projection onto a fixed direction



Projection onto a fixed direction

$$\sigma^2_{X_v} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} \Sigma_X \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$= 229$$

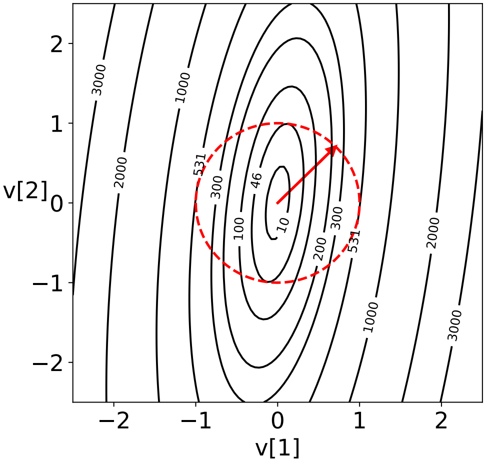
Sample variance = 229 (sample std = 15.1)



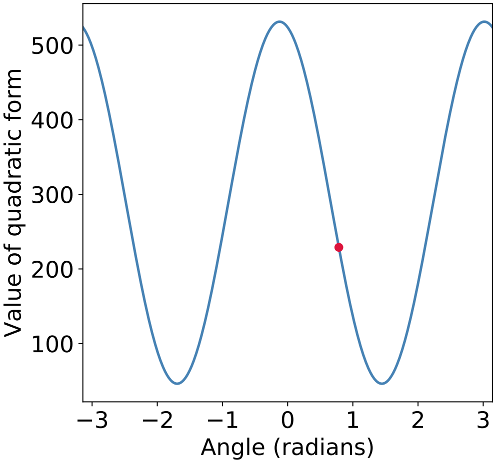
Projection onto a fixed direction

$$\sigma_{X_v}^2 = v^T \Sigma_X v$$

$f(v) := v^T \Sigma_X v$ for $||v||_2 = 1$



$f(v) := v^T \Sigma_X v$ for $||v||_2 = 1$



What have we learned

Covariance matrix encodes variance of any linear combination of a random vector