NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

NYU DATA SCIENCE

Principal component analysis

DS GA 1002 Probability and Statistics for Data Science

Carlos Fernandez-Granda

Prerequisites

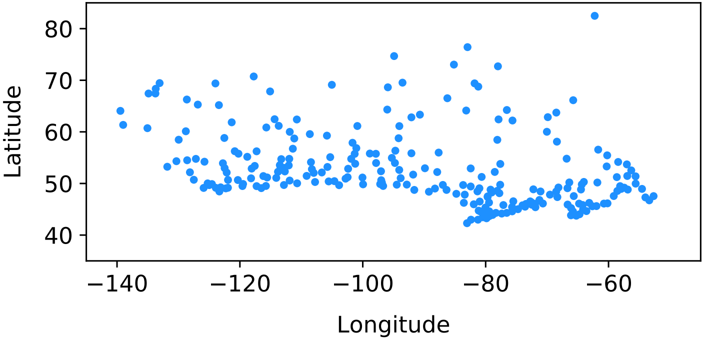
Linear algebra (vectors, inner products, projections, eigendecomposition)

Covariance matrix

Goals

Describe principal component analysis

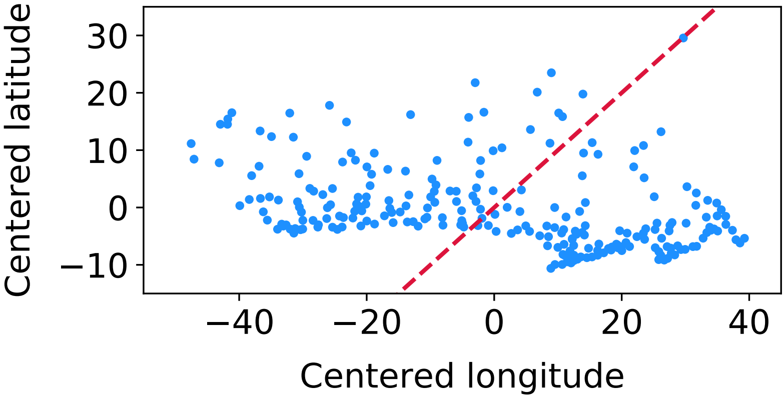
Cities in Canada



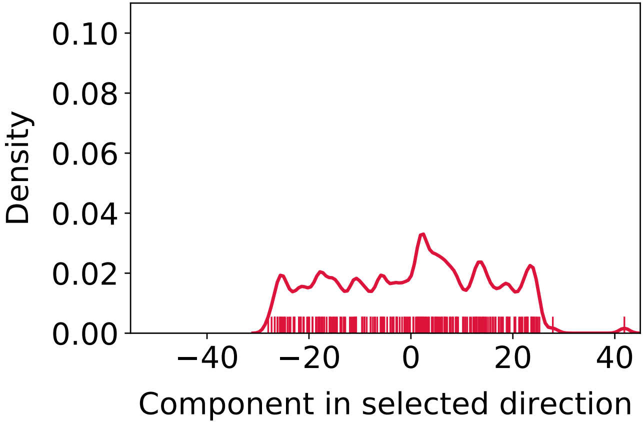
Sample covariance matrix:

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

Projection onto a fixed direction



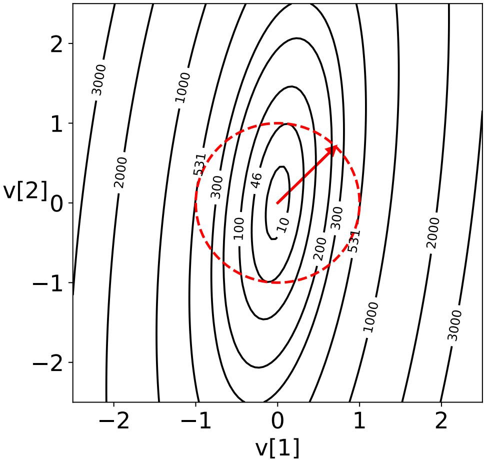
Projection onto a fixed direction



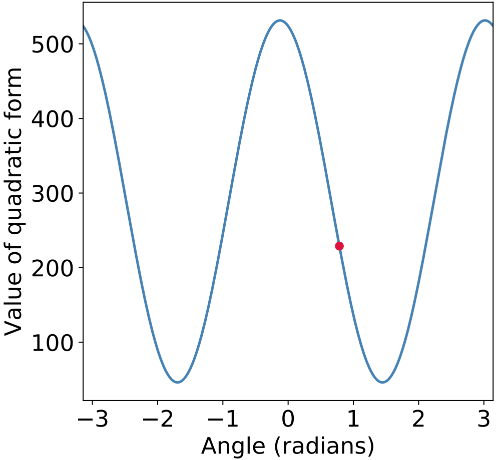
Projection onto a fixed direction

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

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



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



How can we find directions of maximum/minimum variance?

Spectral theorem

If $A \in \mathbb{R}^{d\times d}$ is symmetric, then it has an eigendecomposition

$$A = \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix}^T,$$

Eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ are real

Eigenvectors $u_1, u_2, \ldots, u_n$ are real and orthogonal

Spectral theorem

$$\lambda_1 = \max_{||x||_2=1} x^T A x$$

$$u_1 = \arg\max_{||x||_2=1} x^T A x$$

$$\lambda_k = \max_{||x||_2=1, x \perp u_1, \ldots, u_{k-1}} x^T A x, \quad 2 \leq k \leq d$$

$$u_k = \arg\max_{||x||_2=1, x \perp u_1, \ldots, u_{k-1}} x^T A x, \quad 2 \leq k \leq d$$

Variance in direction of a fixed vector v

If random vector $\tilde{x}$ has covariance matrix $\Sigma_{\tilde{x}}$

$$\mathrm{Var}\left(v^T \tilde{x}\right) = v^T \Sigma_{\tilde{x}} v$$

Principal directions

Let u_1, ..., u_d, and $\lambda_1 > \ldots > \lambda_d$ be the eigenvectors/eigenvalues of $\Sigma_{\tilde{x}}$

$$\lambda_1 = \max_{||v||_2=1} \mathrm{Var}(v^T \tilde{x})$$

$$u_1 = \arg \max_{||v||_2=1} \mathrm{Var}(v^T \tilde{x})$$

$$\lambda_k = \max_{||v||_2=1, v \perp u_1, \ldots, u_{k-1}} \mathrm{Var}(v^T \tilde{x}), \quad 2 \leq k \leq d$$

$$u_k = \arg \max_{||v||_2=1, v \perp u_1, \ldots, u_{k-1}} \mathrm{Var}(v^T \tilde{x}), \quad 2 \leq k \leq d$$

Principal components

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

$$\widetilde{pc}[i] := u_i^T c(\tilde{x}), \quad 1 \leq i \leq d$$

is the ith principal component

$$\mathrm{Var}(\widetilde{pc}[i]) := \lambda_i, \quad 1 \leq i \leq d$$

Principal components are uncorrelated

$$
\begin{aligned}
\mathrm{E}(\widetilde{pc}[i]\widetilde{pc}[j]) &= \mathrm{E}(u_i^T c(\tilde{x}) u_j^T c(\tilde{x})) \\
&= u_i^T \mathrm{E}(c(\tilde{x}) c(\tilde{x})^T) u_j \\
&= u_i^T \Sigma_{\tilde{x}} u_j \\
&= \lambda_j u_i^T u_j \\
&= 0
\end{aligned}
$$

Principal components

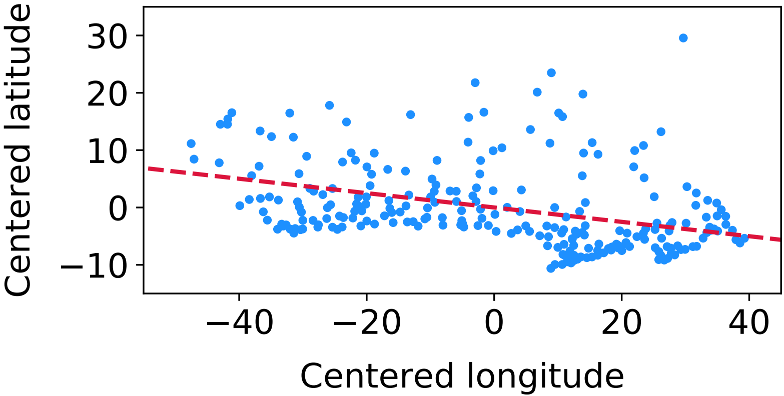
For dataset X containing $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$

1. Compute sample covariance matrix Σ_X
2. Eigendecomposition of Σ_X yields principal directions u_1, ..., u_d
3. Center the data and compute principal components

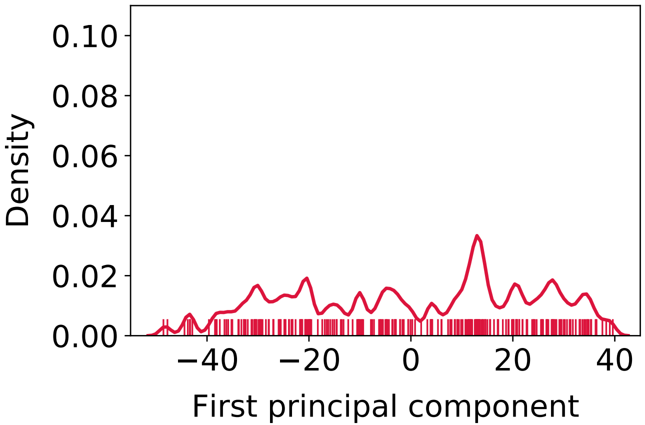
$$pc_i[j] := u_j^T c(x_i), \quad 1 \leq i \leq n, \ 1 \leq j \leq d,$$

where $c(x_i) := x_i - \mu_X$

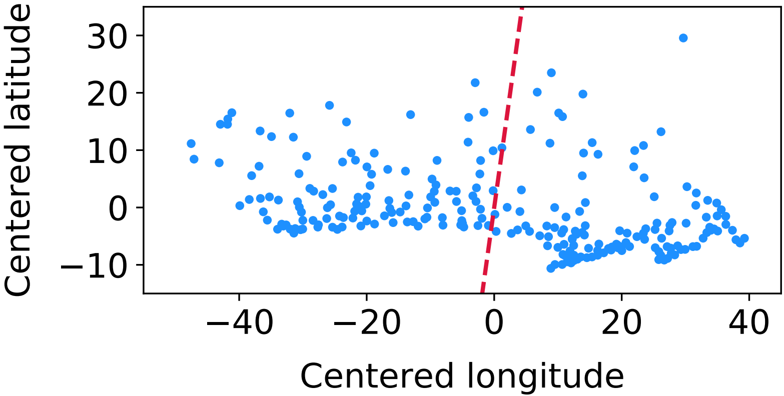
First principal direction



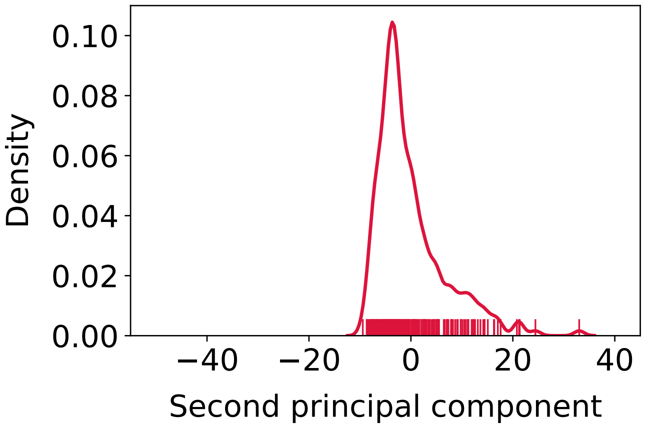
First principal component



Second principal direction



Second principal component



Sample variance in direction of a fixed vector v

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

Principal directions

Let u_1, ..., u_d, and $\lambda_1 > \ldots > \lambda_d$ be the eigenvectors/eigenvalues of Σ_X

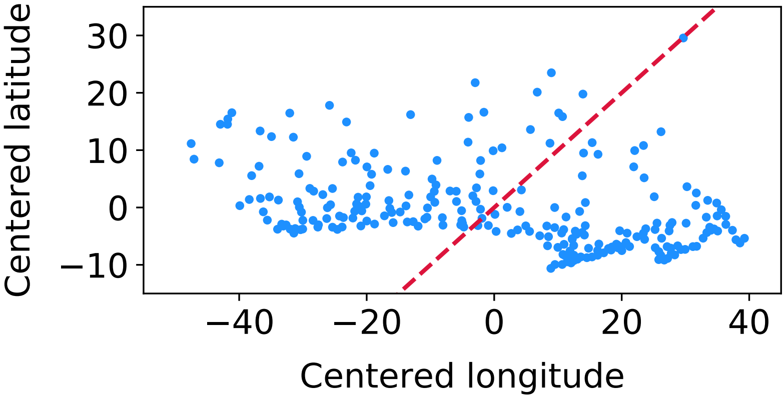
$$\lambda_1 = \max_{||v||_2=1} \sigma^2_{X_v}$$

$$u_1 = \arg\max_{||v||_2=1} \sigma^2_{X_v}$$

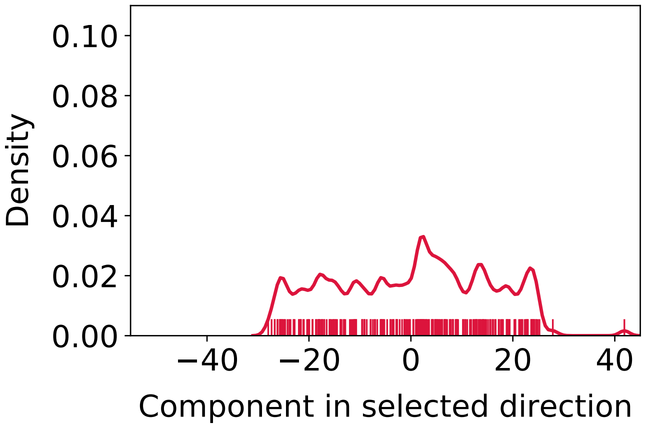
$$\lambda_k = \max_{||v||_2=1, v \perp u_1, \ldots, u_{k-1}} \sigma^2_{X_v}, \quad 2 \leq k \leq d$$

$$u_k = \arg\max_{||v||_2=1, v \perp u_1, \ldots, u_{k-1}} \sigma^2_{X_v}, \quad 2 \leq k \leq d$$

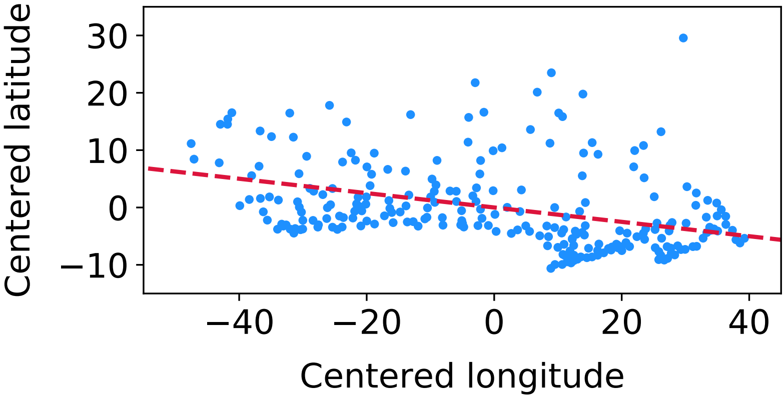
Sample variance = 229 (sample std = 15.1)



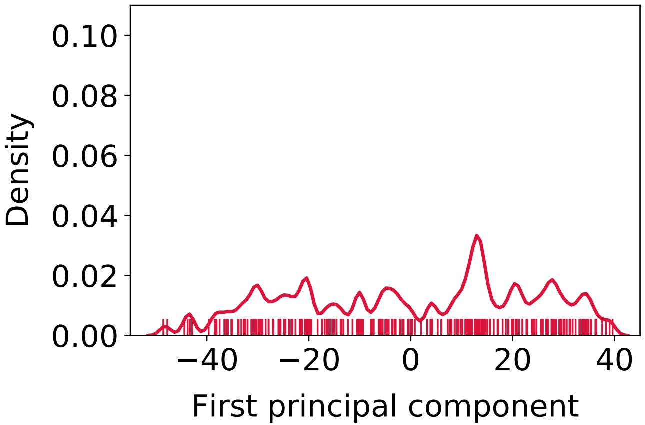
Sample variance = 229 (sample std = 15.1)



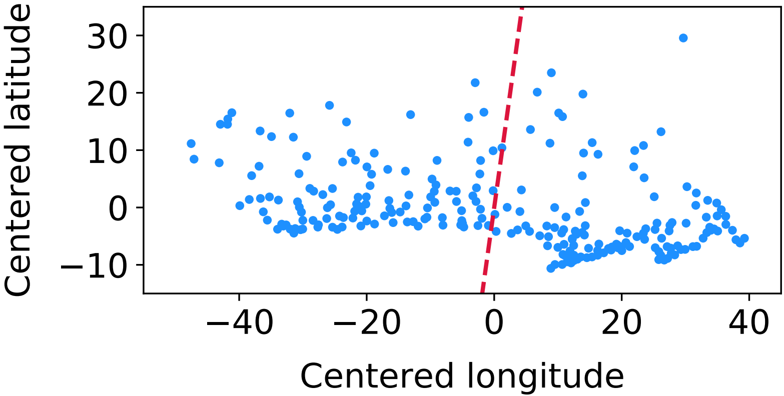
Sample variance = 531 (sample std = 23.1)



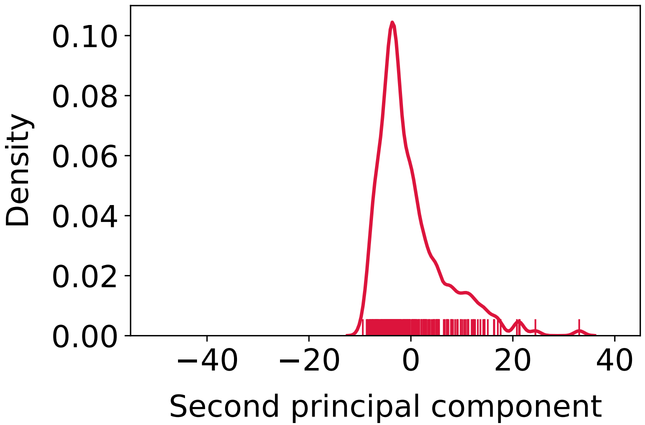
Sample variance = 531 (sample std = 23.1



Sample variance = 46.2 (sample std = 6.80)



Sample variance = 46.2 (sample std = 6.80)

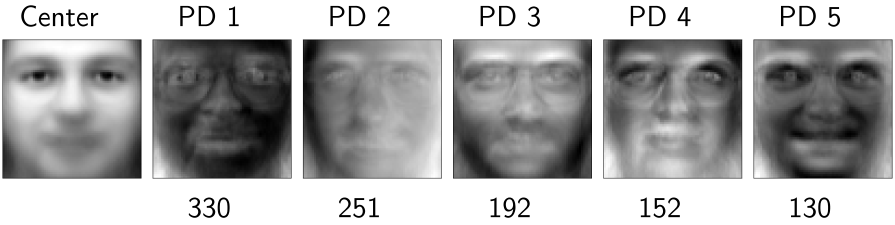


PCA of faces

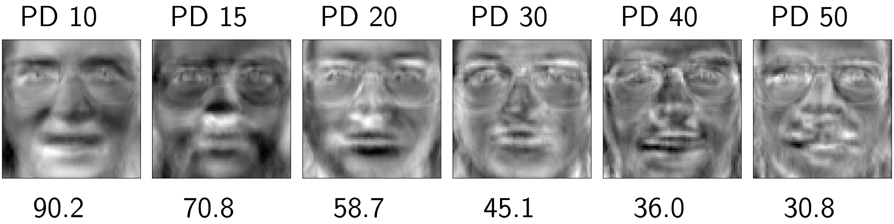
Data set of 400 64×64 images from 40 subjects (10 per subject)

Each face is vectorized and interpreted as a vector in $\mathbb{R}^{4096}$

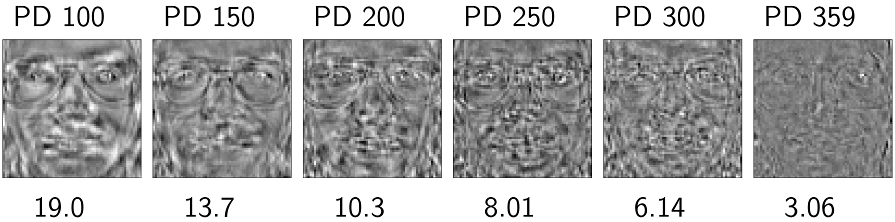
PCA of faces



PCA of faces



PCA of faces



What have we learned

Principal component analysis extracts directions of maximum/minimum variance in the data