

# ENM 360: Introduction to Data-driven Modeling

## *Lecture #27: Principal component analysis*

Paris Perdikaris  
December 8, 2020



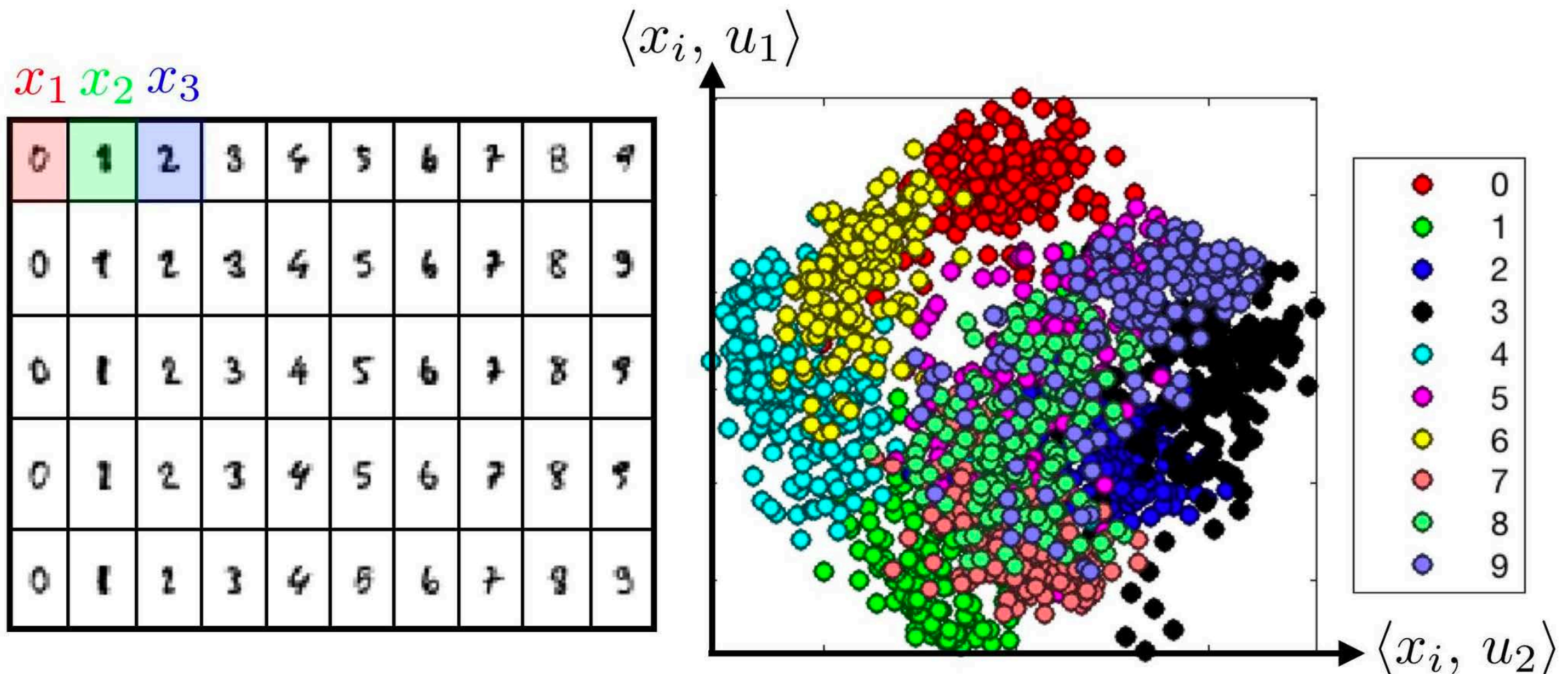
# Principal component analysis

Input data:  $X = (x_i)_{i=1}^n \in \mathbb{R}^{n \times p}, x_i \in \mathbb{R}^p$

Remove mean:  $x_i \leftarrow x_i - \frac{1}{n} \sum_j x_j$

Covariance:  $C \stackrel{\text{def.}}{=} \frac{1}{n} X^\top X \in \mathbb{R}^{p \times p}$

Eigen-decomposition:  $C = U \text{diag}(\sigma_k^2) U^\top, U = (u_k)_{k=1}^p$



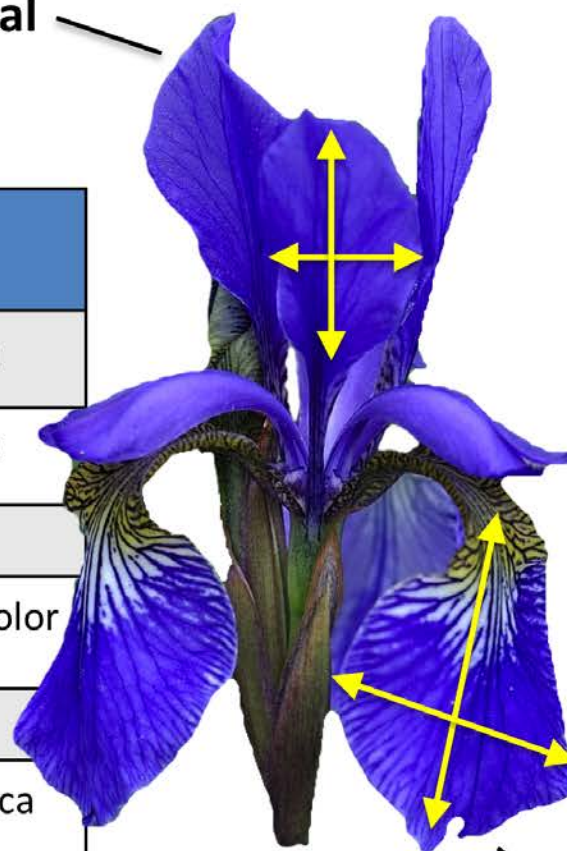
# Example: the Iris data set

## Samples

(instances, observations)

|     | Sepal length | Sepal width | Petal length | Petal width | Class label |
|-----|--------------|-------------|--------------|-------------|-------------|
| 1   | 5.1          | 3.5         | 1.4          | 0.2         | Setosa      |
| 2   | 4.9          | 3.0         | 1.4          | 0.2         | Setosa      |
| ... |              |             |              |             |             |
| 50  | 6.4          | 3.5         | 4.5          | 1.2         | Versicolor  |
| ... |              |             |              |             |             |
| 150 | 5.9          | 3.0         | 5.0          | 1.8         | Virginica   |

Petal

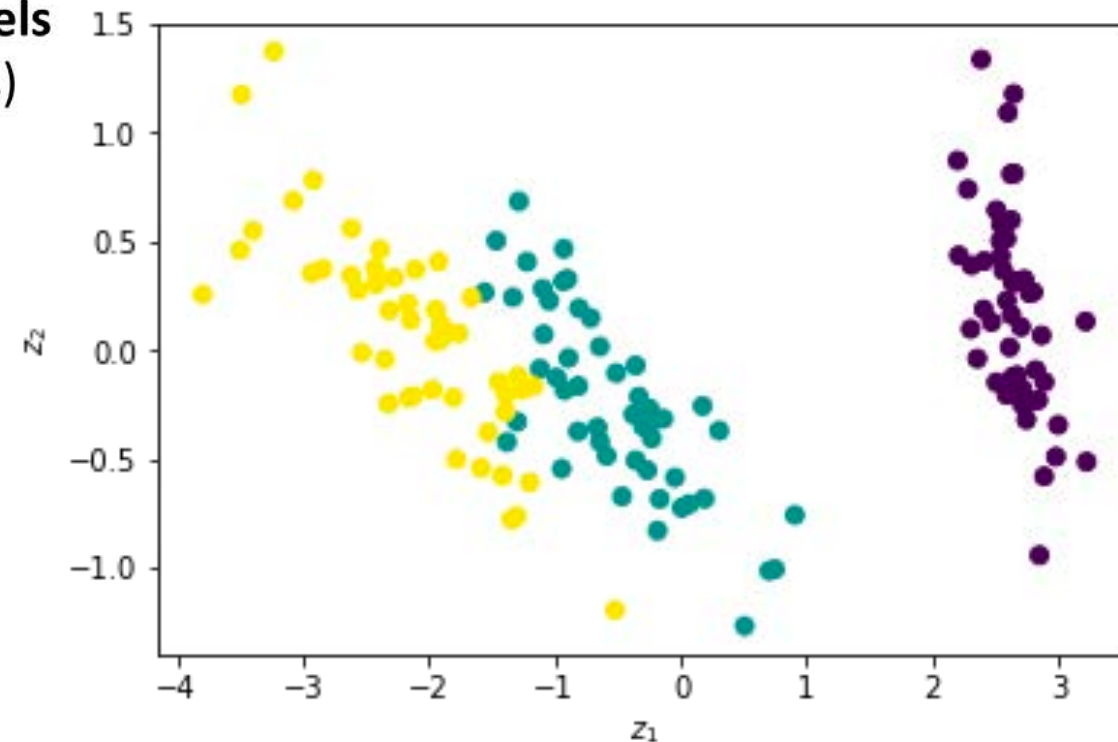


Sepal

Class labels  
(targets)

## Features

(attributes, measurements, dimensions)





# Example: Eigenfaces

## PCA example: Eigen Faces

input: dataset of  $N$  face images

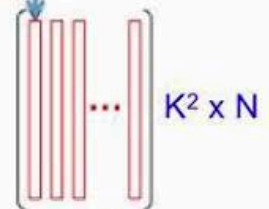


face:  $K \times K$  bitmap of pixels



"unfold" each bitmap to  $K^2$ -dimensional vector

arrange in a matrix  
each face = column



"fold" into a  $K \times K$  bitmap

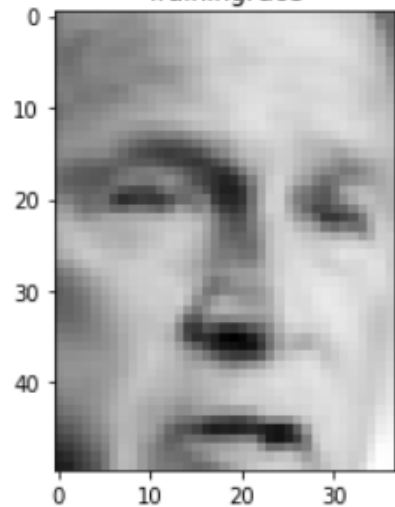


PCA

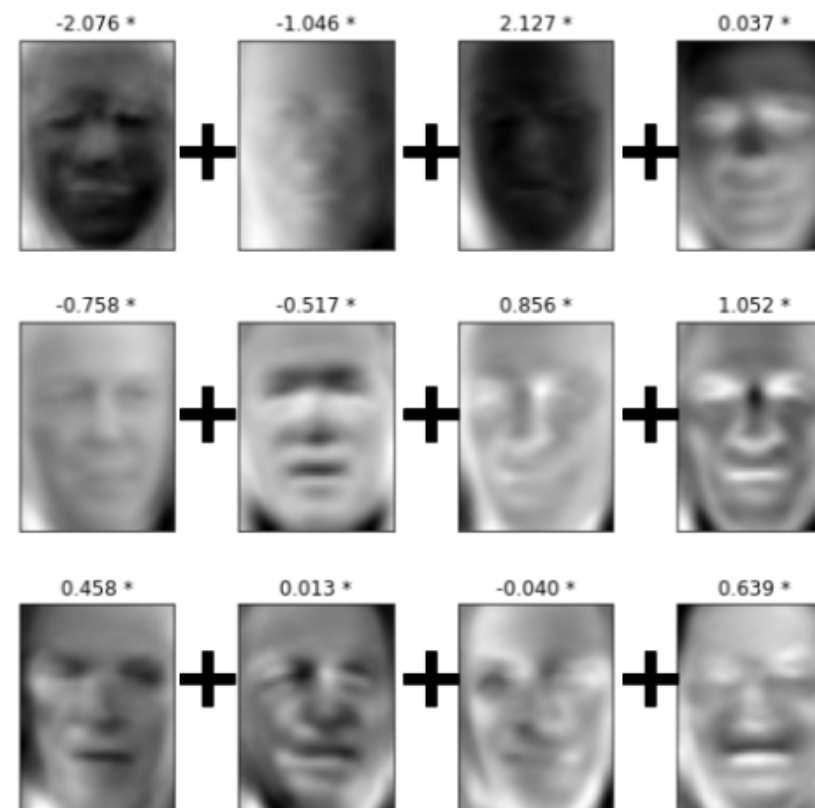
$K^2 \times m$

set of  $m$  eigenvectors  
each is  $K^2$ -dimensional

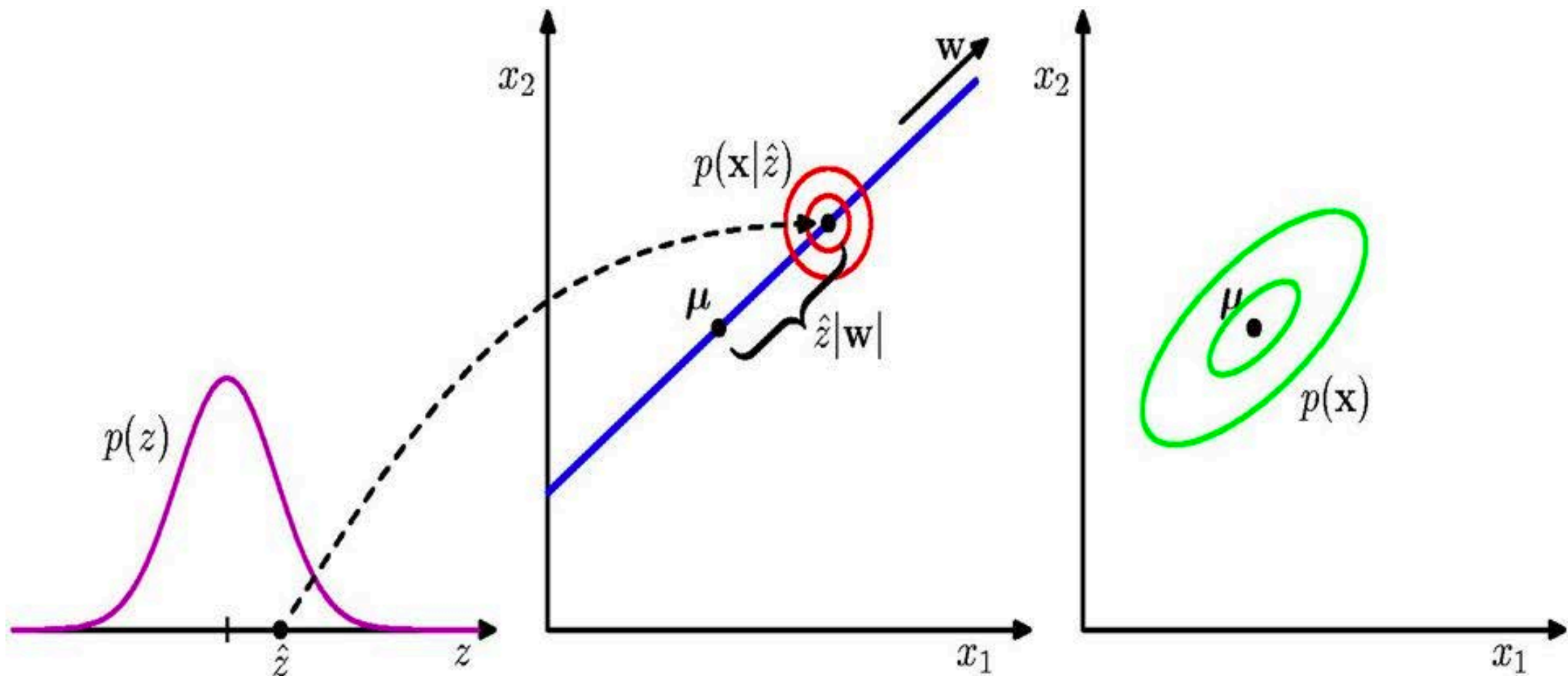
TrainingFace



=



# Example: Probabilistic PCA



data space: 2-dimensional      latent space: 1-dimensional

- get a value  $\hat{z}$  for the latent variable  $z$
- get a value for  $\mathbf{x}$  from an isotropic Gaussian distribution

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \mu, \sigma^2\mathbf{I})$$

- green ellipses: density contours for the marginal distribution  $p(\mathbf{x})$