INTRODUCTION

Classification

Problem setting

Classification methods subdivide data into several, distinct classes. More formally:

- Data $x_1, x_2, ...$
- Each observations falls into one of K categories (the *classes*).
- Learning task: Find a classification function

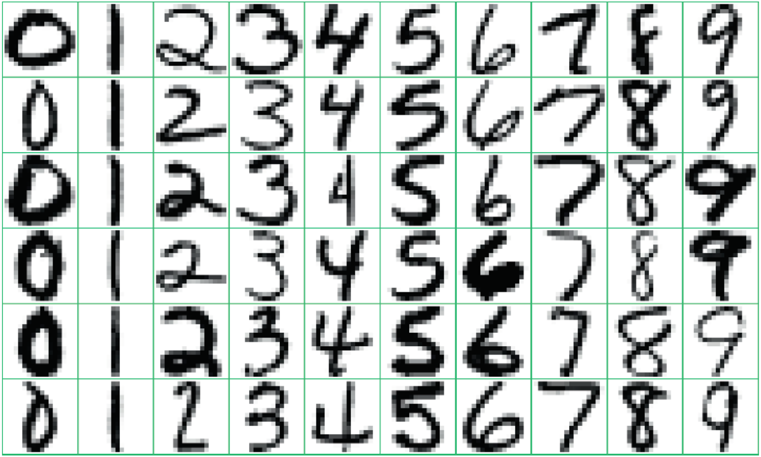
$$f : \mathbf{X} \to \{1, \dots, K\} \;.$$

- Input of the learning problem: Correctly categorized examples $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n$.

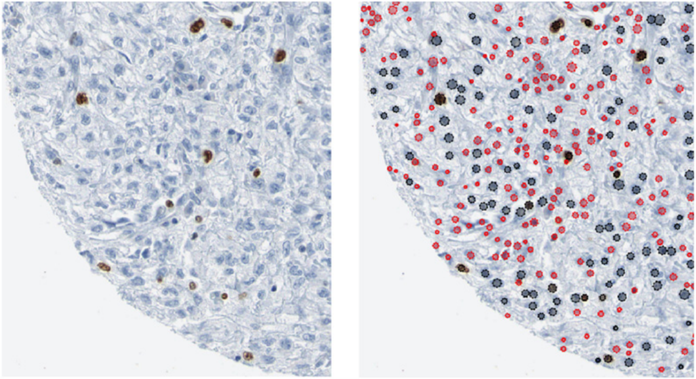
Approach

- Define:
  1. A set of possible classification functions f (the *hypothesis set*).
  2. A *cost function* which assumes a large value when mistakes are made.
- To find a good classifier, search the hypothesis class for the f which keeps costs as small as possible.
- Different types of errors can be more or less expensive.

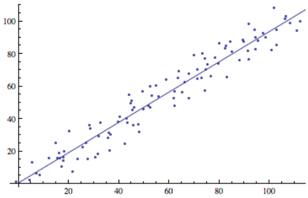
USPS DATA



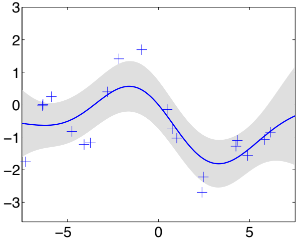
Each digit: 16×16 pixels, i.e. $x \in \mathbb{R}^{256}$



Linear regression



Nonlinear regression



Shrinkage

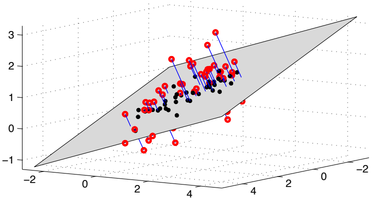
Shrinkage methods make regression less susceptible to idiosyncracies of a specific data set.

- ℓ_2 shrinkage: Reduce variability of solutions between data sets.
- ℓ_1 shrinkage: Weed out unimportant dimensions in high-dimensional data.

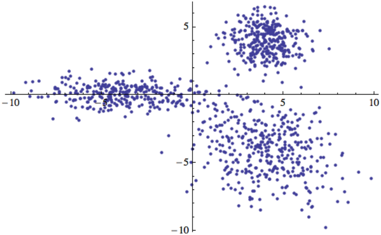
Task

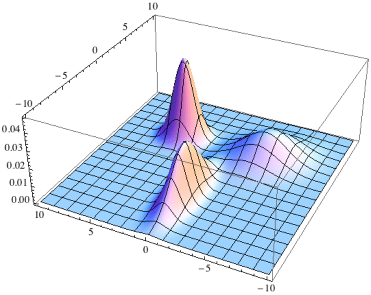
- Input: A high-dimensional data set.
- Output: A low-dimensional representation of the data *which preserves important structure*.

Illustration: From 3D to 2D



CLUSTERING





Example: Image Segmentation