FITTING MODELS to DATA

Oct 18th 2021

Parametric models

Decision boundaries

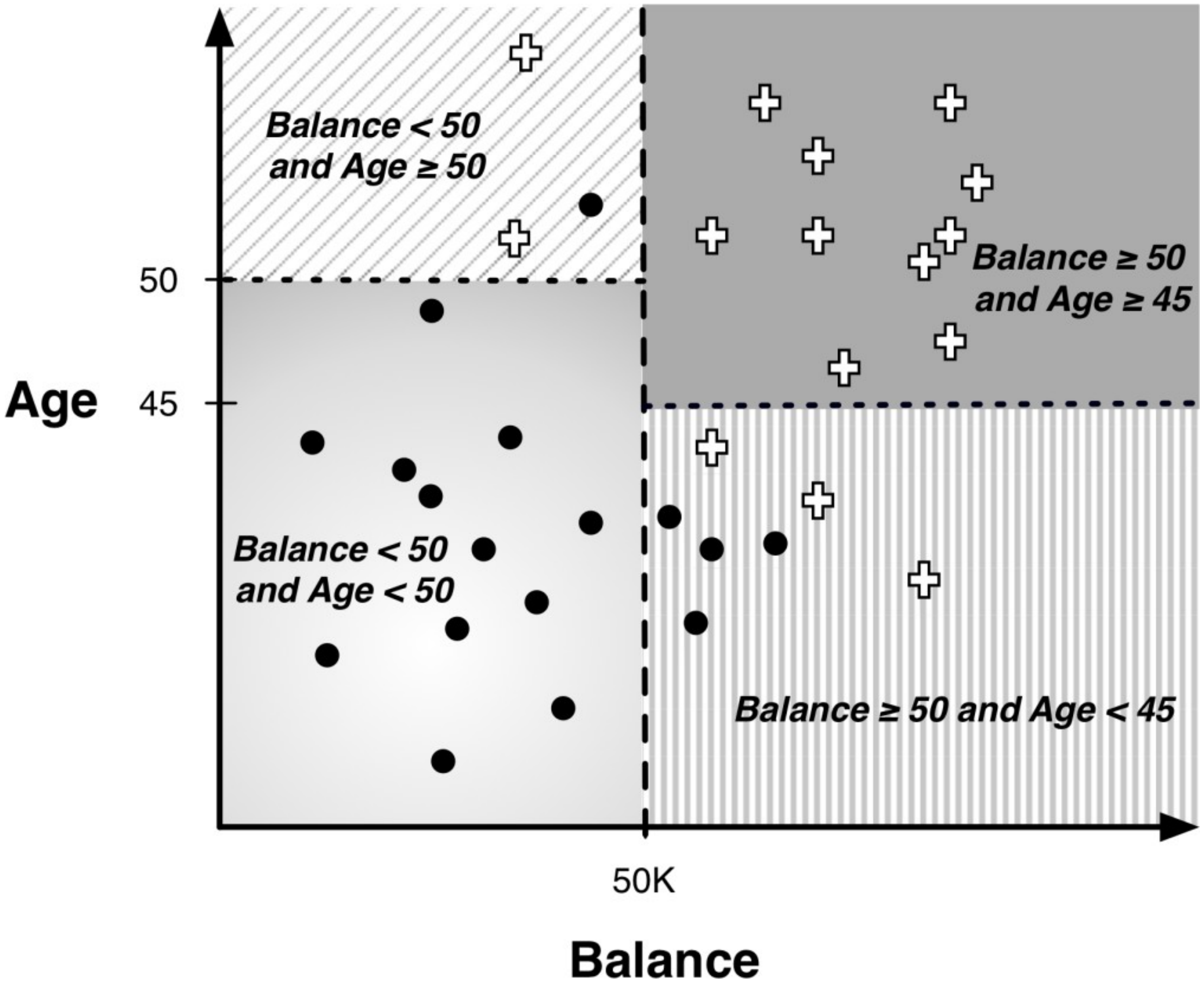
The Linear classifier

Loss and error

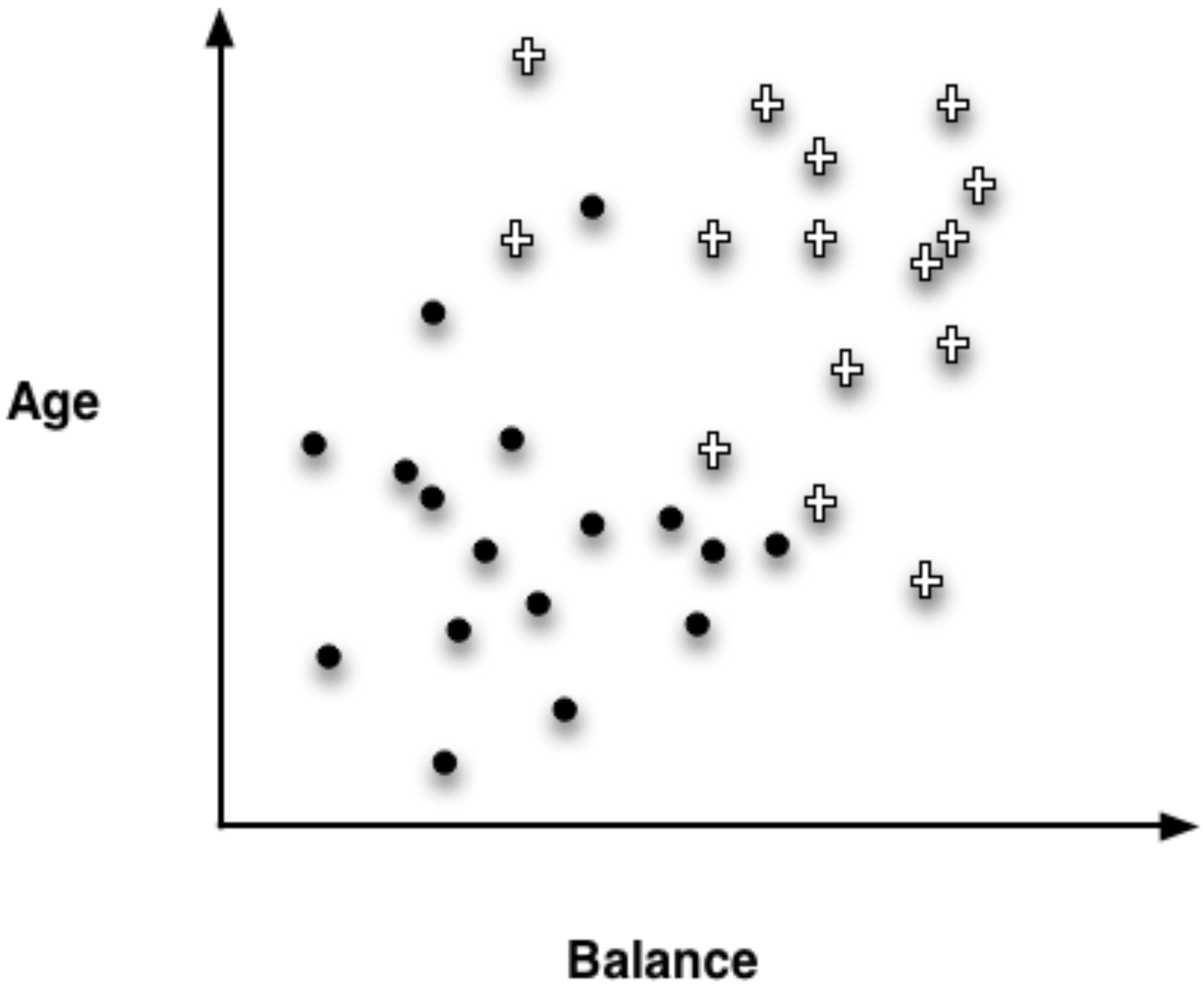
Parametric Models

- Mathematical function
- Parametric models
- The LINEAR MODEL

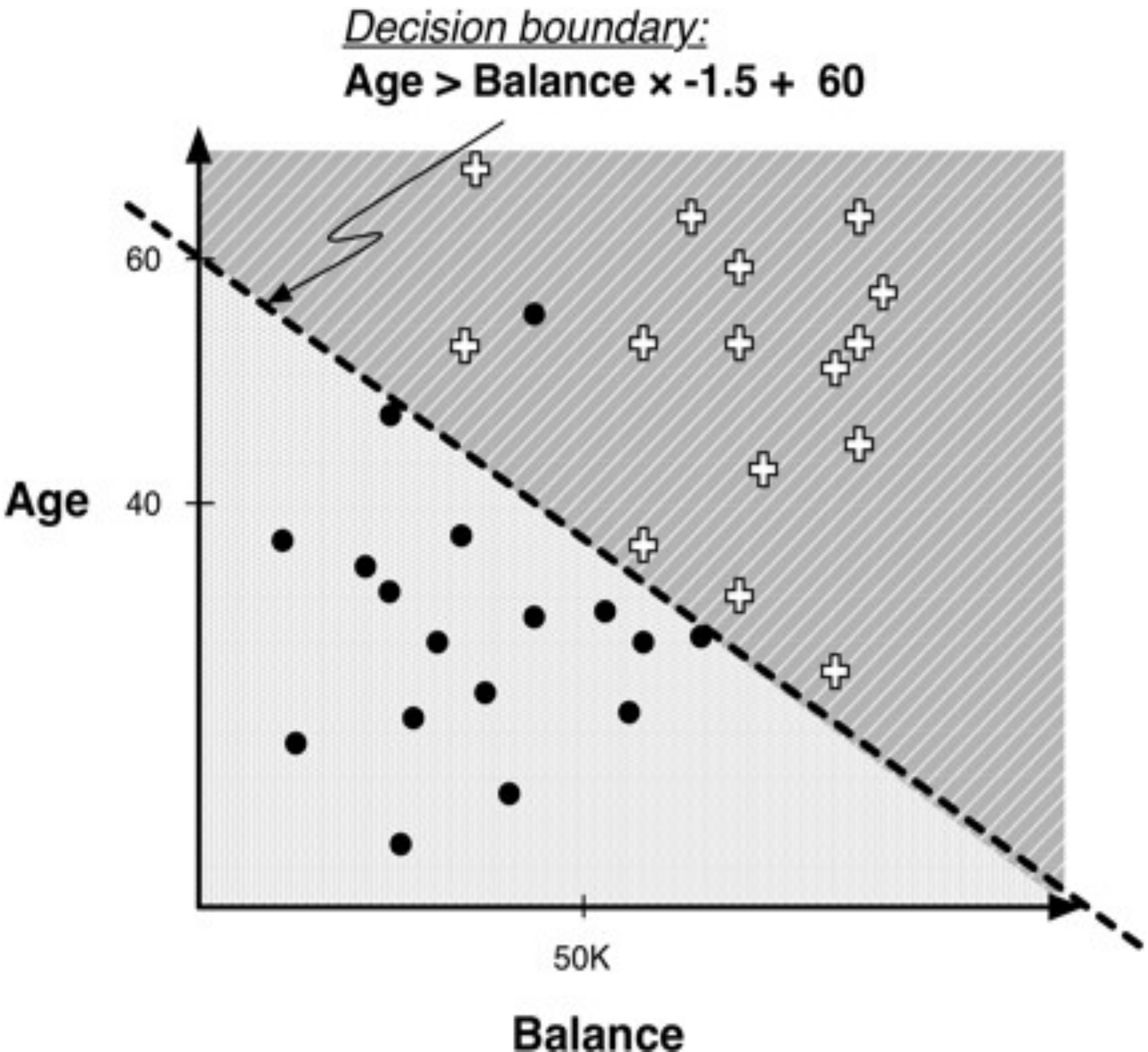
Decision Boundaries



Instance Space



Linear Classifier



Example of Classification Function

- Linear discriminant:

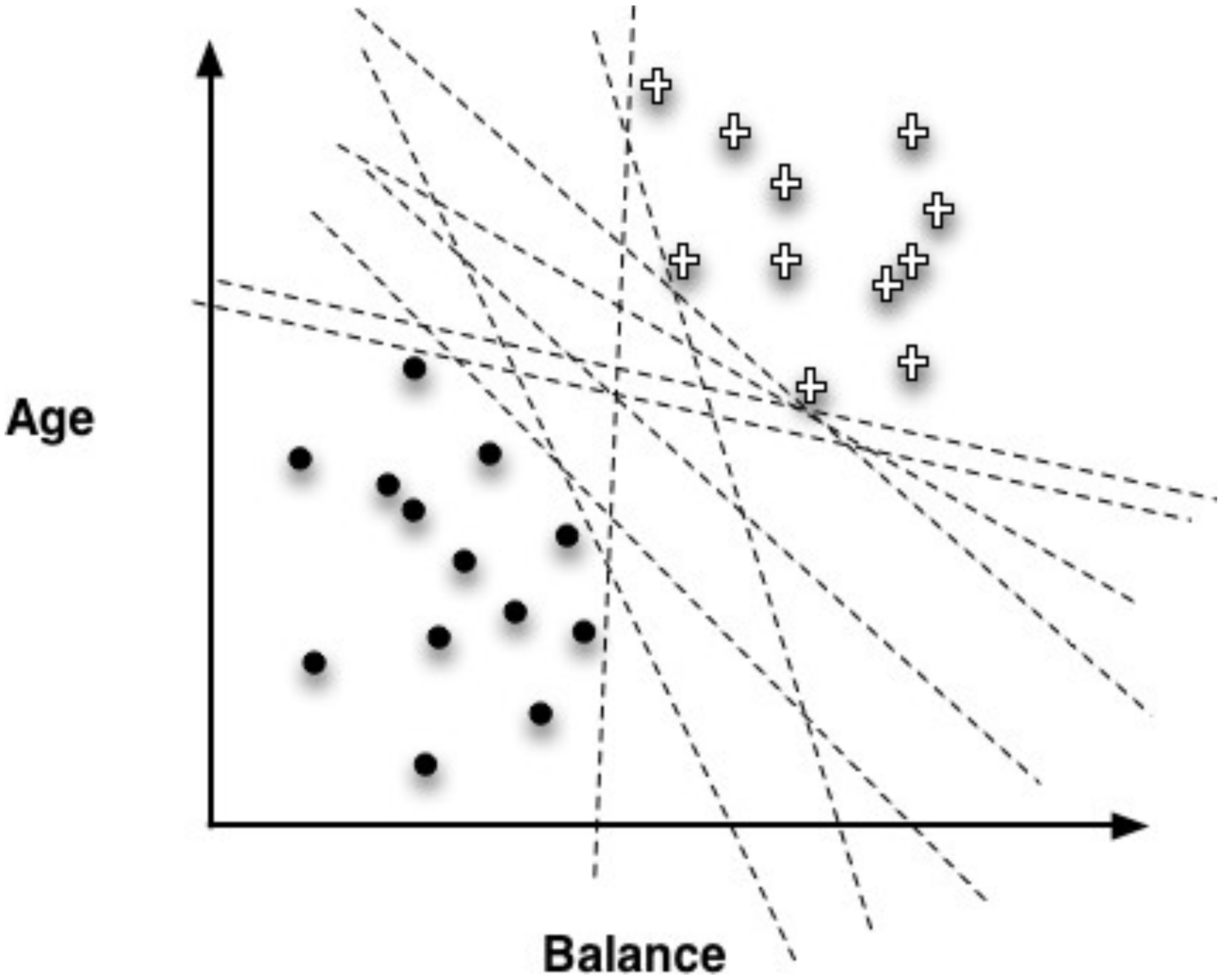
$$Class(x) = \begin{cases} + & \text{if } 1.0 \times Age - 1.5 \times Balance + 60 > 0 \\ \bullet & \text{if } 1.0 \times Age - 1.5 \times Balance + 60 \leq 0 \end{cases}$$

- We now have a **parameterized model**: the weights of the linear function are the parameters

- The weights are often *loosely* interpreted as **importance indicators** of the features

- A different sort of multivariate supervised segmentation
  - The difference from DTs is that the method for taking multiple attributes into account is to create a mathematical function of them

Choosing the "best" line



Generalized Linear Model

$$\boldsymbol{f(x) = w_o + w_1 x_1 + w_1 x_1 + \ \ldots}$$

Objective Functions

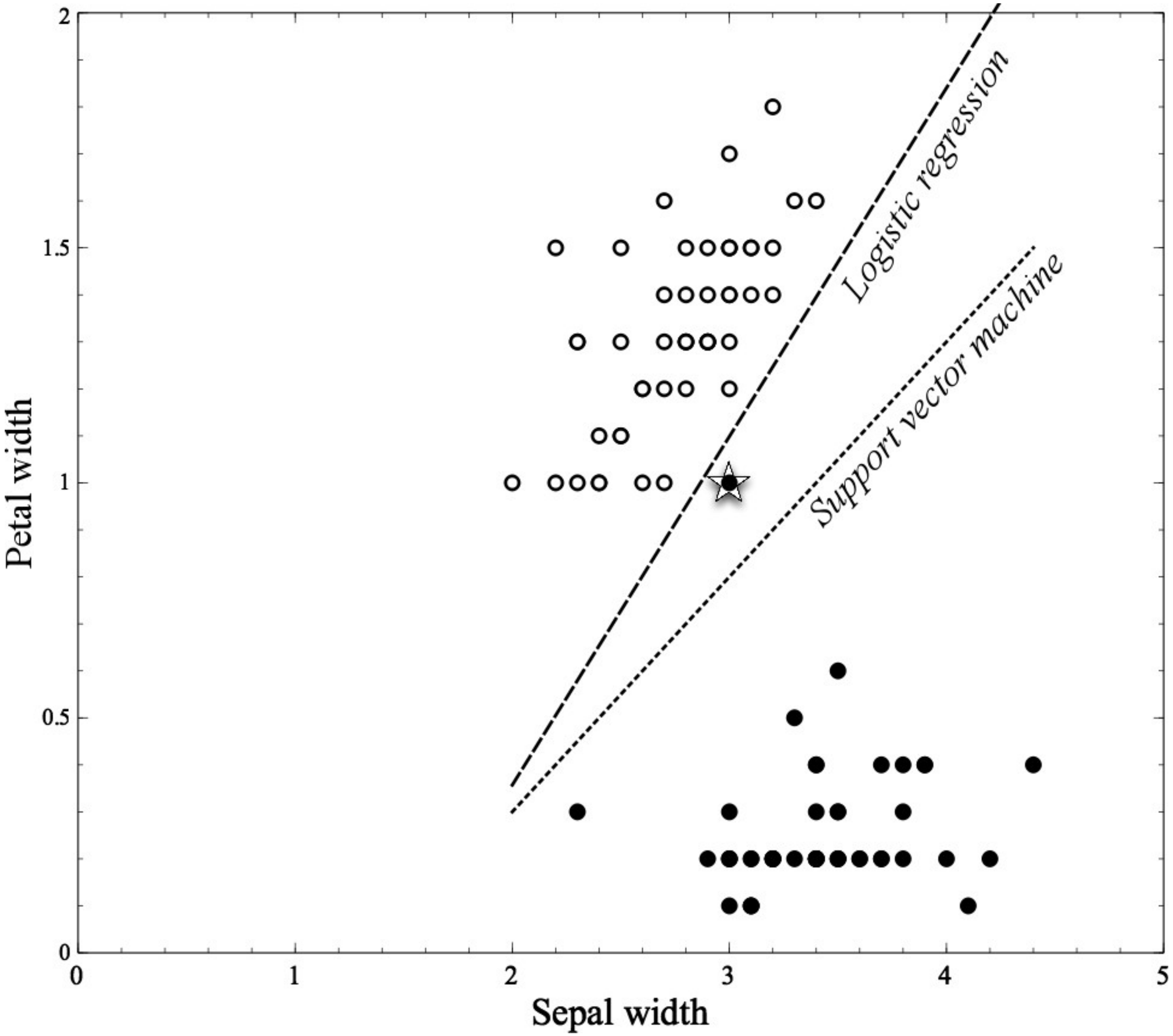
- "Best" line depends on the **objective (loss) function**
  - Objective function should represent our goal

- A loss function determines how much penalty should be assigned to an instance based on the error in the model's predicted value

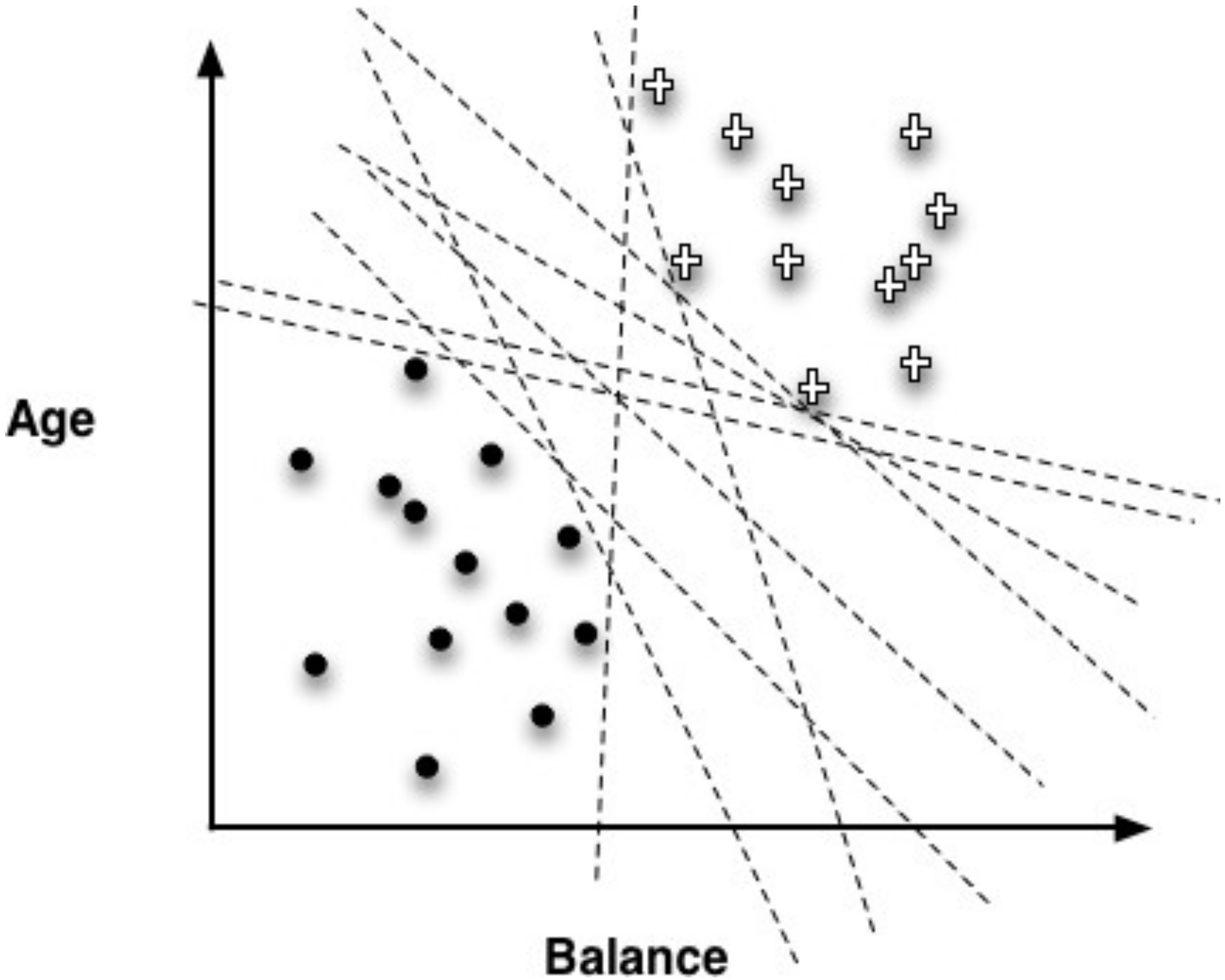
- Examples of objective (or loss) functions:
  - $\lambda(y;x) = |y - f(x)|$
  - $\lambda(y;x) = (y - f(x))^2$ [convenient mathematically – linear regression]
  - $\lambda(y;x) = I(y \neq f(x))$

- **Linear regression**, **logistic regression**, and **support vector machines** are all very similar instances of our basic fundamental technique:
  - The key difference is that each uses **a different objective function**

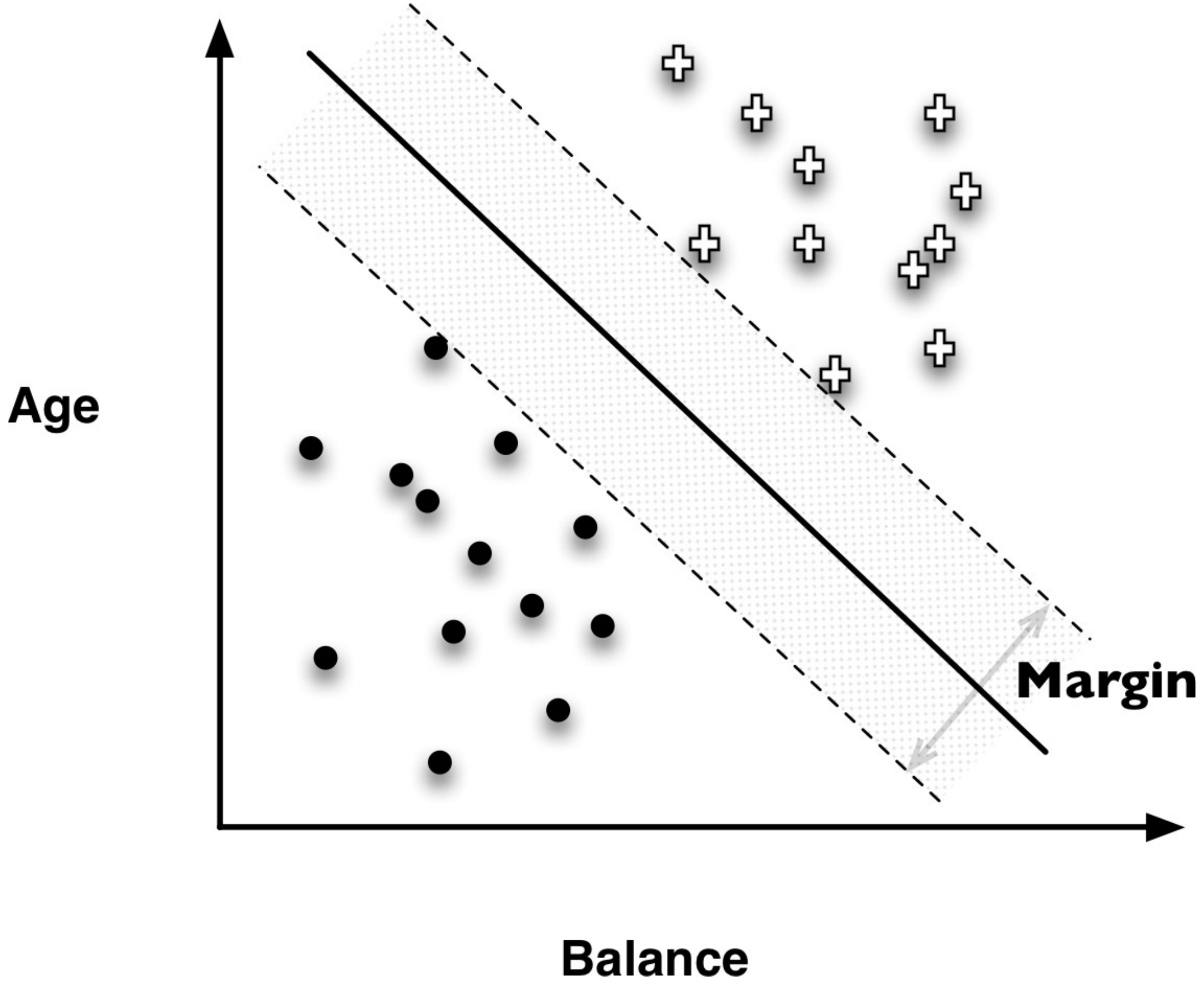
Classifying Flowers



Choosing the “best” line



Support Vector Machines (SVMs)

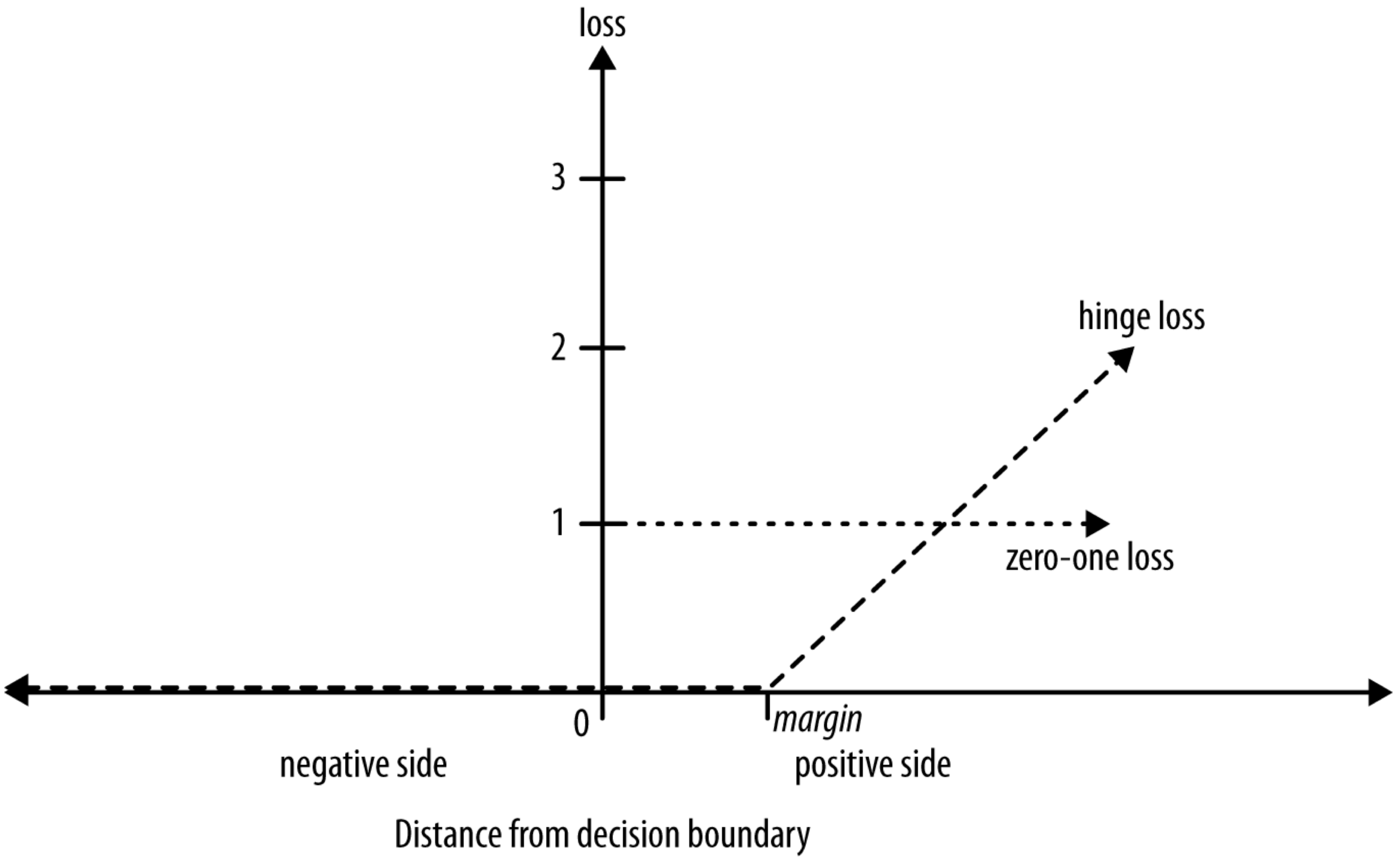


Hinge Loss Functions

- Support vector machines use **hinge loss**
- Hinge loss incurs no penalty for an example that is <u>not</u> on the wrong side of the margin
- The hinge loss only becomes positive when an example is on the wrong side of the boundary and beyond the margin
  - Loss then increases linearly with the example's distance from the margin
  - Penalizes points more the farther they are from the separating boundary

Loss Functions

- **Zero-one loss** assigns a loss of zero for a correct decision and one for an incorrect decision
- **Squared error** specifies a loss proportional to the square of the distance from the boundary
  - Squared error loss usually is used for numeric value prediction (regression), rather than classification
  - The squaring of the error has the effect of greatly penalizing predictions that are grossly wrong


loss
3
2
1
hinge loss
zero-one loss
0
margin
negative side
positive side
Distance from decision boundary

Error

- Absolute error
- Least squares
  - Sum or average of

Time for Lab !