Machine learning

Wenda Zhou

June 20, 2017

Machine learning

- Machine learning can be seen as an alternate paradigm for understanding data.
- Combine ideas of computer science, statistics and applied math.
- Mostly interested in various prediction problems

Prediction in machine learning

Usually have dataset of observations (y_i, x_i), and wish to understand the relation:

$$y = f(x) \tag{1}$$

Prediction in machine learning

Usually have dataset of observations (y_i, x_i), and wish to understand the relation:

$$y = f(x) \tag{1}$$

Statistics Interested in some aspect of f, e.g. does f depend on x_1.

Machine learning Interested in obtaining the "best" f

Example 0: random forest

The function f is given by a random forest.

- Good predictive performance
- General method
- Difficult to interpret predictions

Example 1: k-NN

Non-parametric algorithm (as data grows, so does the model).

k-nearest neighbours for classification

Look at k closest observations in training set. Predict most likely class.

Example 1: k-NN

Non-parametric algorithm (as data grows, so does the model).

k-nearest neighbours for classification

Look at k closest observations in training set. Predict most likely class.

k-nearest neighbours for regression

Look at k closest observations in training set. Predict average value.

Example 1: k-NN

Non-parametric algorithm (as data grows, so does the model).

k-nearest neighbours for classification

Look at k closest observations in training set. Predict most likely class.

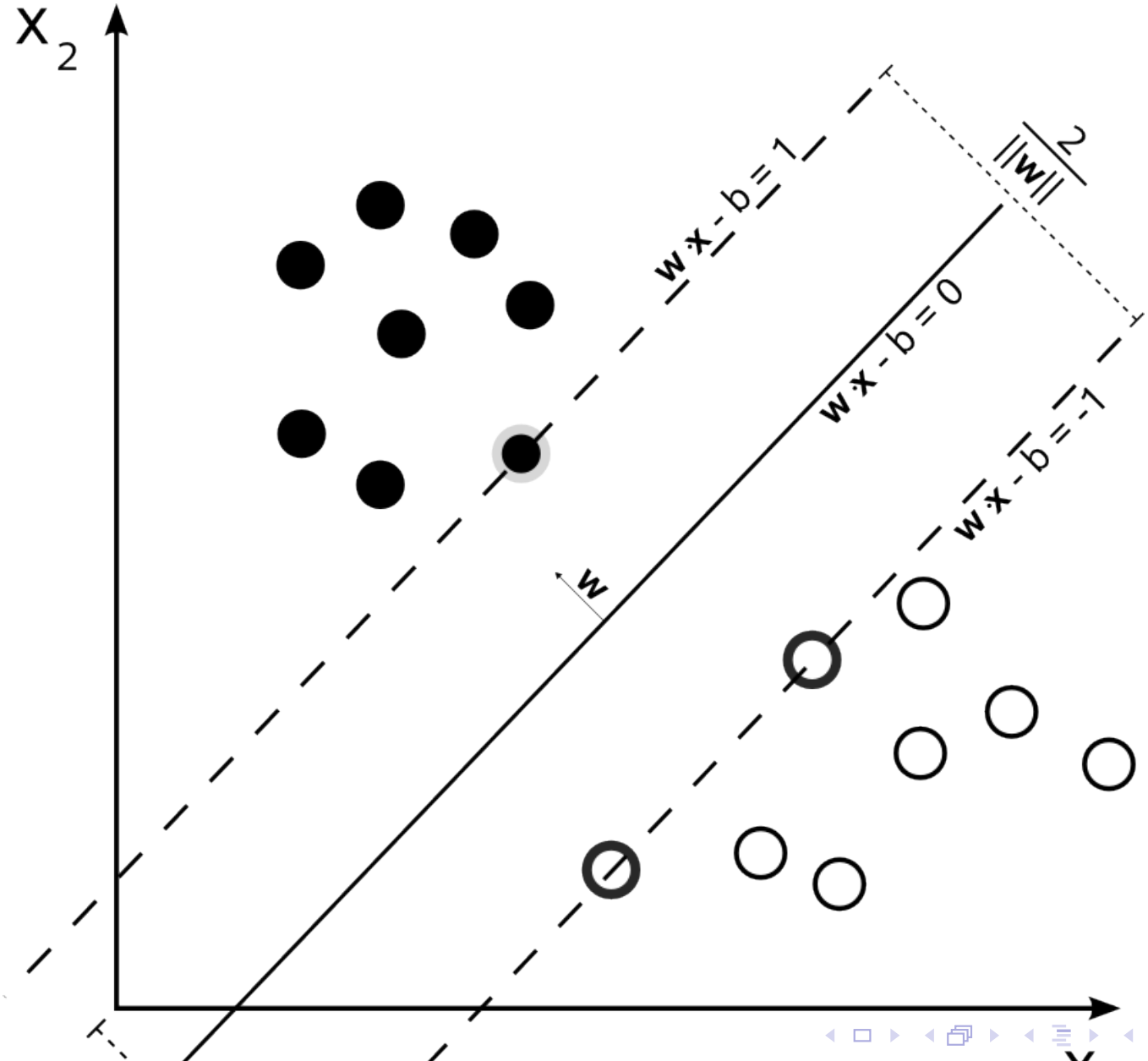
k-nearest neighbours for regression

Look at k closest observations in training set. Predict average value.

Tuning parameter k creates bias-variance trade-off.

Example 2: SVM

Large margin linear classifier.



Example 2: SVM

- Large margin linear classifier.
- Can also classify non-linear boundaries by using the "kernel" trick.
- Similar to how we can use linear regression to fit polynomial models.

Example 3: neural networks

- Neural networks can be seen as stacking several linear models together with a non-linearity between them.
- Best model for prediction – especially when large amounts of data available.
- Can be adapted to most problems that have large amounts of available data: image, text, speech etc.

Gradient descent

From an abstract point of view, can write the neural network as a function f_A parametrised by A (millions of parameters in modern applications).
Find parameters A that minimize the loss.

$$L(A) = \sum_{i=1}^{n} (y_i - f_A(x_i))^2 \tag{2}$$

Gradient descent

From an abstract point of view, can write the neural network as a function f_A parametrised by A (millions of parameters in modern applications).
Find parameters A that minimize the loss.

$$L(A) = \sum_{i=1}^{n} (y_i - f_A(x_i))^2 \tag{2}$$

Gradient descent

Proceed iteratively, at each step, go in the direction of the gradient (derivative) to minimize objective.

Conclusion

- Machine learning methods are particularly adapted for prediction
- Requires large amounts of data but can be extremely effective
- Predictions may be dificult to interpret