Lecture 2

Supervised Learning I:

kNN:

* Simple and effective if we have a distance measure that captures the  
  semantic meaning of the task.

kNN and distance measures:

* Euclidean
* Manhatten
* Chebyshev
* Jaccard

Supervised learning:

Supervised learning involves observing several examples of an arbitrary  
⃗𝑥 together with an associated label or vector ⃗𝑦 from a feature space.  
Then learning to predict ⃗ 𝑦 from ⃗ 𝑥.

Given a training set of ℕ example input-output pairs  
(𝑥1, 𝑦1), (𝑥2, 𝑦2), ⋯, (𝑥n, 𝑦n) where each 𝑦i was generated by an  
unknown function 𝑦i = 𝑓 (𝑥i), we want to discover a function ℎ that  
approximates the true function 𝑓.

Hypothesis space

* The hypothesis space is the set of all functions ℎ that approximates  
  the true function 𝑓.
* A learning problem is realizable if the hypothesis space contains the  
  true function 𝑓.
* We cannot always tell if a learing problem is realizable
* We go for approximations!

Finding a good hypothesis

* Learning is a search through the space of possible hypothesis to find  
  one that performs well.
* We may measure how well a hypothesis performs in terms of  
  accuracy. To do this we “give it” a test set of examples that are  
  distinct from from the training set.
* Accuracy – the fraction of examples from the test set for which the  
  output was assigned correctly.

Representation

* Scalar
* Vector
* Matrix
* Tensor

Scalar

* Single number or value

Vector

* A vector is arranged in order.
* We can identify each number by its index  
  ⃗𝑥 = [𝑥1 , 𝑥2 , ⋯ , 𝑥n]   
  Beware, there are many notations for vectors.

Matrix

* A matrix is a two-dimensional array of numbers.
* We can identify each number in a matrix by two indices.

Tensor

* A tensor is an array of numbers arranged on a regular grid with a  
  variable number of axes.
* An n-ranked tensor has n indices

Linear algebra

* Algebra: The study of (mathematical) structures and the rules for  
  manipulating them
* Linear: 𝑎1𝑥1 + ⋯ + 𝑎n𝑥n = 𝑏

Curse of dimensionality

As the number of features increase, our data covers less of the total  
space. Increasing the risk of overfitting, and we need to collect more  
examples to avoid the curse.