

Parametric models

A particular model class is assumed (e.g. linear)

• Number of parameters fixed in advance

Note: even if you K-fold CV to try different candidate models with different number of parameters, we are still defining each candidate model in advance

Non-parametric models

• Minimal assumptions about model class • Model structure determined by data

1-NN - runtime

• Training: just store data • Inference: need to:

– compute distance to each point

– distance metric typically scales with d

The runtime for predicting each test point is $O(d)$

What value of K ?

• In general: larger K , less complex model

How does this compare to linear regression or logistic regression?

There, the training time could be long, but inference was $O(1)$.

• K often cited "rule of thumb": use $K = \sqrt{N}$

We can tolerate a long training time more easily than a long inference time.

* Often cited "rule of thumb": use $K = \text{sqrt}(N)$

Note that there will be zero error on training set (unless there are training data points that have identical feature values, but different labels).

K nearest neighbors

Instead of 1 closest sample, we find K . Let N_k be the set of K training points that are closest to x_0 .

How do we use this set for (1) classification? (2) regression?

Figure 3: 3-nearest neighbors.

KNN for classification

Idea: Estimate conditional probability for a class as fraction of points among neighbors with the class label.

Remember: Let N_k be the set of K training points that are closest to x_0 .

Then, we can estimate the per-class conditional probability given the sample x_0 :

For each class $m \in M$:

$$P(m|x_0) = \frac{1}{K} \sum_{x_i \in N_k} I(y_i = m)$$

where $I(y_i = m)$ is 1 if $(x_i, y_i) \in N_k$ is a member of class m , 0 otherwise.

• We can then select the class with the highest probability.

• Practically: select the most frequent class among the neighbors.

KNN for regression

Idea: Use the combined label of the K nearest neighbors. For example, we can take their mean:

$$\hat{y}_0 = \frac{1}{K} \sum_{x_i \in N_k} y_i$$

Figure 4: Example of a regression using median vote of the 3NN. The "true function" is shown in black, the orange dashed line shows the prediction of the regression model.

Distance Functions

(L2 distance prefers many medium-sized disagreements to one big one.)

There are many more choices - for example, look at the distance metrics implemented in sklearn.

Problems with the basic distance metric:

• When features have different scale/range, need to standardize

• KNN implicitly weights all features equally: this is a problem if you have features that are not relevant for the target variable

• For images: pixel-wise distance doesn't necessarily equate to perceptual similarity

With L1 distance, we had a data-driven way to do feature selection.

The nearest neighbor method doesn't have any "built-in" way to do feature weighting or feature selection as part of the training process.

so we need to do it ourselves as part of the pre-processing steps.

Bias and variance of KNN

The function

Suppose data has true relation

$$y = f(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

and our model predicts $\hat{y} = f(x)$.

Assumption of training set

For this derivation, we consider the expectation over:

- the test points
- the error
- the randomness in the y values in the training set

We do not consider randomness in the x values - we assume a fixed training set.

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

• the test points are drawn from the same distribution as the training set

• the randomness in the y values is the same as the training set

• the x values are the same as the test points

