

Notes on Machine learning

Debdeep Bhattacharya

January 14, 2019

1 Documentaion

- scikit-learn tutorial

2 numpy matrix operations

Create a new vector with `np.array([1, 2, 3])` Create a new matrix with `np.c_[1, 2, 3]`
Transpose with `.T` Create a random matrix with entries from a normal distribution with size 2x1 using `np.random.normal(size = (2,1))`

3 k-Nearest Neighbor (k-NN)

If the number of features (dependent variables) is p and the maximum distance of any two points in a cluster is less than d , then the required data points (sample size) need to be $n \sim \frac{1}{d^p}$.

4 Python

Install scikit-learn on linux using `pip3` (not `pip`) and run with `python3`, not `python`.