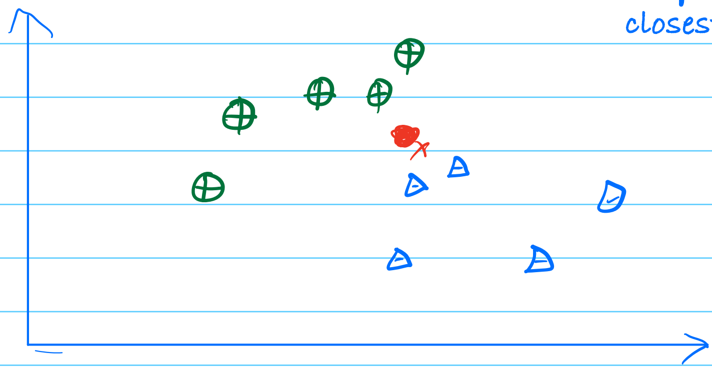
CS 3/5780

K-Nearest Neighbors Classifier

Assumption: Similar points are likely to share same label

Classification Rule: For a test point x, assign it label that is the most common label amongst the k $\underbrace{\text{most similar}}_{\text{closest}}$ training instances.



Formally: Given $D = \{(x_1,y_1), (x_2,y_2),\dots,x_n,y_n)\}$ and test point x

Find $S_x \subseteq D$ s.t. $|S_x| = k$ and $\forall\ (x',y') \in D \setminus S_x$

$$dist(x',x) \geq \max_{(x'',y'')\in S_x} dist(x'',x) \qquad h(x) = MODE(y' : (x',y') \in S_x)$$

Pro-tip: In case of a tie, reduce k by 1 and repeat

Common distances: Minkowski distance or lp distances

$$dist(x, x') = \left(\sum_{i=1}^{d} |x[i] - x'[i]|^p\right)^{1/p}$$

$p = 2$ is Euclidean distance

Quiz: What distances do $p=0$, $p=1$ and $p \to \infty$ correspond to

Bayes Optimal Classifier

If we knew P how well could we do? $\qquad (x,y) \sim P$

Bayes Error: $\min_{\text{All possible } h} E[\ell(h(x),y)]$ $(= P(h(x) \neq y))$

If you knew $P(Y \mid X=x)$, at Given point x, optimal classifier:

$$h_{opt}(x) = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \; P(Y=y \mid X=x)$$

$$\text{Bayes Error}(x) = 1 - \max_{y \in \mathcal{Y}} P(Y=y \mid X=x)$$

This is the Best we can do!

Quiz: say a coin has probability p of falling heads when tossed

1. If the coin is tossed twice, what is the probability of two different out comes?
2. Show that this probability is lesser than $2(1-p)$

1 - NN classifier : simplest case $k=1$

Risk of 1-NN $\leq$ 2 Bayes ERROR

Formal proof is involved, see Cover & Hart '67

Intuition:

1. Say P was a discrete distribution on a finite set of points. Then, as $n \to \infty$ every test. point has already occurred in DTRAIN. (say we pick any one of previous occurrence as the nearest neighbor
2. Risk of 1-NN Classifier is now given by the quiz Question. Why?

We are asking the question, what is the probability that, label y of a new test instance x matches that of a randomly chosen training point $x_i \in D_{Train}$ such that, $x_i = x$. Its label y_i is drawn independently from $P(Y \mid X=x)$

Hence, $P(h_{NN}(x) \neq y \mid X=x) \leq 2\left(1 - \max_{y \in \{0,1\}} P(Y=y \mid X=x)\right)$

Claim: Given x, let $\hat{x} \in D_{TRAIN}$ be the 1-NN of x in D_{TRAIN} as $n \to$

$$\text{dist}(x, \hat{x}) \to 0, \quad x \to \hat{x}$$

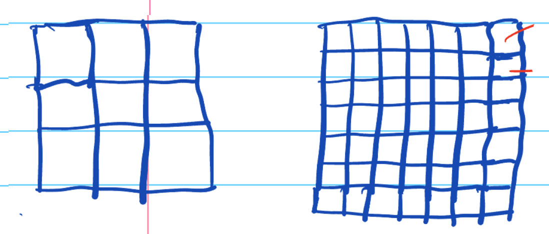
Risk of 1-NN ≤ 2 Bayes ERROR

k-NN for general, $k > 1$

a. use larger k as n becomes larger.

Intuition: For point x, we predict as majority of k draws from $P(Y|X=x)$

b. But, if k grows too fast, the, more, we rely on farther away points to predict label for x



1. Each sub-cube has same area
2. if we randomly pick a few sub-cubes, we will pick more cubes near surface!

1. Assume points are drawn uniformly at random from a unit-hypercube
2. Hypercube of volume k/n within unit hypercube is expected to contain k out of n points
3. length ℓ of such cube is given by $\ell^d = k/n$

What does this mean?

$k = 10$
$n = 1000$ →

d	2	10	100	1000
ℓ	0.1	0.63	0.955	0.9954

$\ell = \left(\frac{k}{n}\right)^{1/d}$