

k-Nearest Neighbour

Bayes Classifier

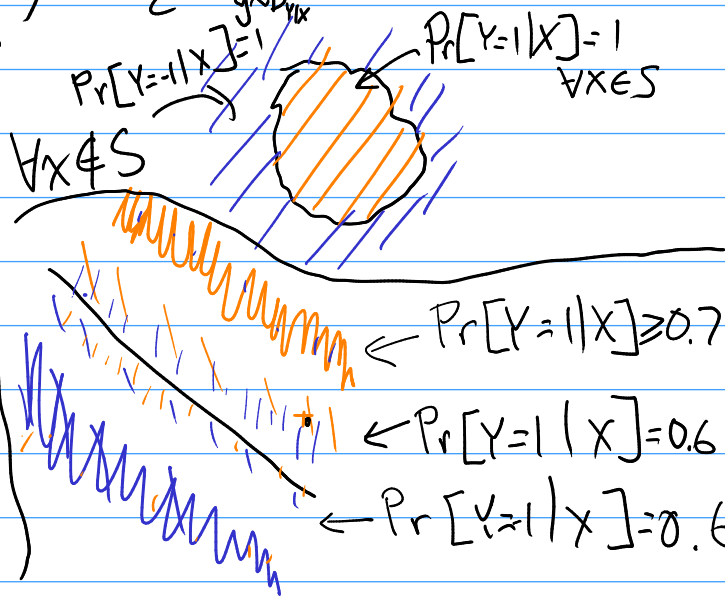
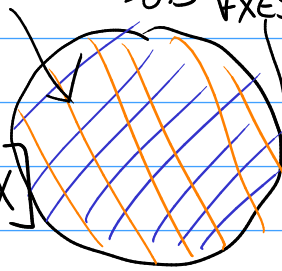
$$\min_{x, y \in D} \Pr[f(x) \neq y]$$

$$f^*(x) = \operatorname{argmax}_c \Pr[y=c | X]$$

$$= \Pr[f^*(x) \neq y]$$

$$= \int_x \Pr[x] \Pr[f^*(x) \neq y | x]$$

$$= \mathbb{E}_{x \sim D_x} \left[1 - \max_c \Pr[y=c | x] \right]$$



k-Nearest Neighbour

$\Pr[y=c | x] \approx \Pr[y=c | x']$ when x and x' are close

Algorithm: kNN

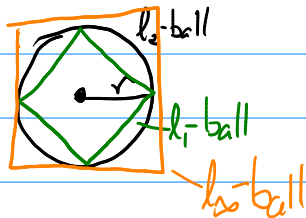
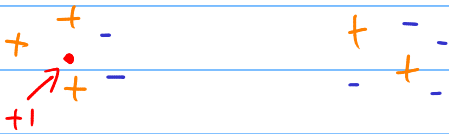
Input: Dataset $D = \{(x_i, y_i) \in X \times Y : i = 1, \dots, n\}$, new instance $x \in X$, hyperparameter k

Output: $y = y(x)$

- 1 for $i = 1, 2, \dots, n$ do
- 2 $d_i \leftarrow \text{dist}(x, x_i)$ // avoid for-loop if possible
- 3 find indices i_1, \dots, i_k of the k smallest entries in d
- 4 $y \leftarrow \text{aggregate}(y_{i_1}, \dots, y_{i_k})$

$k=5$

-non-parametric

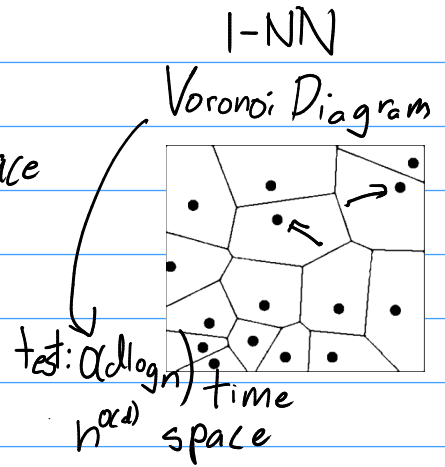


Time + Space Complexity

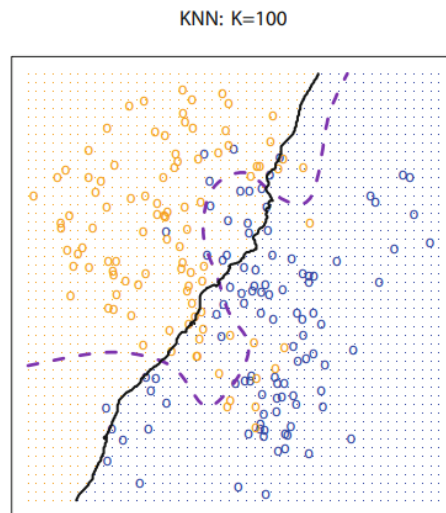
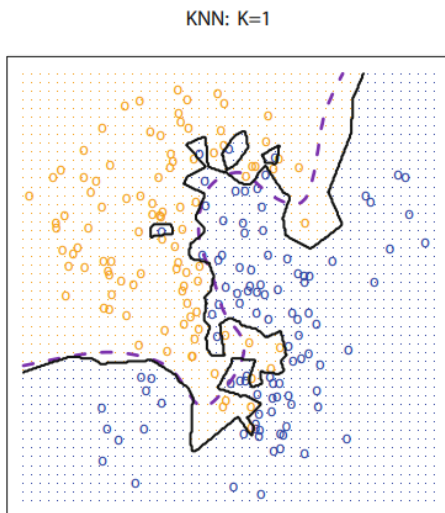
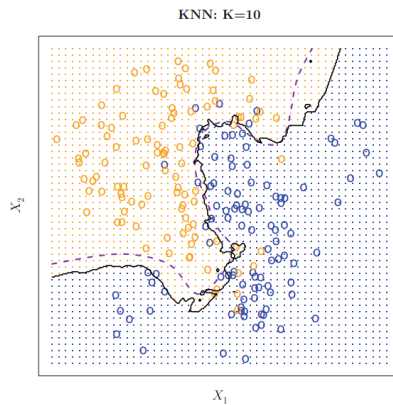
Train: $O(\text{time})$, $O(nd)$ space

Test: $O(ndk)$ time $\rightarrow O(nd)$ time, $O(nd)$ space

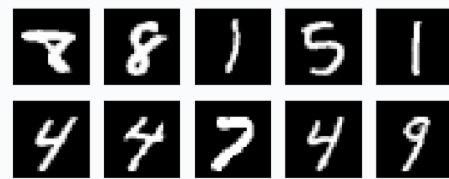
Approx NN



Effect of k



Does it work on MNIST?
 60k train, 10k test $28 \times 28 = 784$



CLASSIFIER	PREPROCESSING	TEST ERROR RATE (%)	Reference
Linear Classifiers			
linear classifier (1-layer NN)	none	12.0	LeCun et al. 1998
K-nearest-neighbors, Euclidean (L2)	none	3.09	Kenneth Wilder, U. Chicago
K-nearest-neighbors, L3	none	2.83	Kenneth Wilder, U. Chicago
K-NN with non-linear deformation (IDM)	shiftable edges	0.54	Keyzers et al. IEEE PAMI 2007
K-NN with non-linear deformation (P2DHMDM)	shiftable edges	0.52	Keyzers et al. IEEE PAMI 2007
2-layer NN, 300 hidden units, mean square error	none	4.7	LeCun et al. 1998
Convolutional net LeNet-4	none	1.1	LeCun et al. 1998

Theory

Suppose $n \rightarrow \infty$. $L_{INN} \leq 2L_{Bayes}(1 - L_{Bayes})$, Cover-Hart 1967

$$\text{If } L_{Bayes} = 0, L_{INN} = 0$$

$$= \frac{1}{2},$$

$$\frac{1}{2}$$

$$= 0.1$$

$$\leq 2 \cdot 0.1 \cdot 0.9$$

$$= 0.18$$

SSBD Curse of dimensionality



\exists a dist. over $[0, 1]^d \times \{\pm 1\}$
 where the Bayes Loss is 0.

But any classifier given $n \leq 1.5^d$ samples
 has error $\geq \frac{1}{4}$.