KNN-2

Machine Learning (ML) 10:00

Agenda:

1. weighted KNN
2. Usecases
3. Bias Variance Tradeoff (M. Imp.)
4. Outliers
5. Interpretability
6. Use case KNN
7. KNN Imputation (Imp.)

"Lazy processor" Methods → ML learning involved.

Train time : $O(1)$

Test time : $O(\text{nlogn} + nd)$

$\Rightarrow O(n \log n)$

Space Complexity → entire data → $O(nd)$

Big eg (row d) → test time ⇒ $\underbrace{(w^T x + b)}_{\text{dot}}$

$\rightarrow O(d)$

→ space complexity ⇒ $(w^T x + b)$

$\underbrace{w}_{d}$ $\underbrace{x}_{d}$ → $O(d+1)$

Weighted KNN

$\boxed{K = 5}$

$\text{pos-vote} = 3$

$\text{neg-vote} = 2$

$w^{(i)} \Rightarrow \frac{1}{d^{(i)}}$

(ex) $d^{(1)} = d^{(2)} = 0.1$ $\quad w^{(1)} = w^{(2)} = 10 = \frac{1}{0.1}$

(ex) $d^{(1)} = d^{(2)} = d^{(3)} = 0.5$ $\quad w^{(1)} = w^{(2)} = w^{(3)} = 2 = \frac{1}{0.5}$

Total weight (−ve class) ⇒ 10+10 ⇒ $\boxed{20}$ (+3 (+ve))

total weight (+ve class) = 2+2+2 → 6

Hyperparameters (KNN)

→ 'homogeneous neighbourhood'

Magic → whether data is homogeneous or not!!!

Distances

→ distance 'hyperparameter'.

→ manhattan dist. (L1)

→ euclidean dist. (L2 dist)

→ minkowski dist (generalized form) hp minm.

Minkowski dist / Lp norm

$$d(x_i, x_j) \Rightarrow \boxed{\left(\sum_{j=1}^{d} |x_{ij} - x_{kj}|^p\right)^{1/p}}$$

$\boxed{p = 2}$ euclidean

$$\sqrt{\sum_{j=1}^{d} (x_{ij} - x_{kj})^2}$$

$\boxed{p = 1}$ manhattan

$$\boxed{\sum_{j=1}^{d} |x_{ij} - x_{kj}|}$$

why euclidean?

data → image → 'Image'.

$A(1,1,1)$ $\quad B(1,1,0)$

$0 + |1| = |1|$

$\Rightarrow 1$

Bias Variance Tradeoff. (Underfitting / overfitting)

$\boxed{K = 1}$ (low)

fitting to outliers

overfit → High Variance low bias

[very complex model]

total datapoints in $\boxed{\text{train}}$ for eval. → Unseen

$\boxed{K = 2/3 \text{ of}}$

'always prefer the class'

→ 'over simplified'

→ Underfit / High Bias

Conclusion:

$K \Uparrow \Rightarrow$ underfit

$K \Downarrow \Rightarrow$ overfit.

How to find best $\boxed{K}$?

err. → $\boxed{\text{cv}}$ → Train → Eval

→ 'cv data'

get best. or diff. 'k value'.

→ underfit

Train Perf. V. Val Perf

Multi-class classification in knn!

'easy'

$\boxed{\text{g NN}}$

$g = -ve$

$g = +ve$ (ok) (ok)

$g = -ve$ (ok)

Outliers

$\boxed{K = 1}$ (low)

↓

affects the confusion matrix

$\boxed{K = 5/7}$

↓

new confusion matrix

K=1, 2, 3, 4, 5, 6, 7, 8, 9

$\boxed{\text{cv/eg}}$ Best, Best, Best, Best, Best, Best, Best, Best, Best.

'Perfect'

Interpretability (KNN)

Big eg → weight

x_q → distance → 'cv/eg'

x_q → $\boxed{\text{cv}}$ → distance → distance.

→ 'Distance'.

'Using brightness for interpretation'

$\boxed{\text{euclidean dist}}$ → Problem → 'it doesn't work in higher dim'

→ Curse dimension → d ↑ ⇒ ↓

x_i is similar to x_j

x_i is not similar to x_j

$\text{dist}(x_i, x_j) \geq \text{dist}(x_i, x_k)$

$\boxed{\text{cosine similarity}}$ (x_i, x_j) ⇒ $\text{Cos}\theta = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$ → $\frac{\text{dot}}{\text{dot}}$

$[-1, 0, 1]$

'Search within d is ↓ ↓ ↓ ↓'

Conclusion:

① (Search) euclidean → (not a high dim)

② Manhattan → always/mostly/sometimes

③ Cosine-sim → better d is high

④ $\boxed{\text{Cosine}}$ $\boxed{\text{distance metric}}$

↑ $\boxed{\text{cosine}}$ $\boxed{\text{(1-cos)}}$

↑ $\boxed{\text{chordal distance}}$

↑ $\boxed{\text{Mahalanobis distance}}$

$\boxed{\text{domain}}$ $\boxed{\text{knowledge}}$

How we get Probabilities from knn instead of labels?

$p(y=1|x) = 0.85$

$\boxed{K = 8}$

5 → +ve

3 → +ve

$P(y=1/x_q) \Rightarrow \frac{5}{8} \Rightarrow 0.62$

$P(y=0/x_q) = \frac{3}{8} \Rightarrow 0.37$

(Q) KNN $\boxed{\text{Multi-class}}$ model!

KNN Imputation → also for missing values

'fill missing values'

→ mean/median

→ mode

x_{10}

$\boxed{\text{Eg}}$

x_{10} missing → $(x_1^{(10)}, x_2^{(10)}, x_3^{(10)}, x_4^{(10)}, x_5^{(10)}, x_6^{(10)}, x_7^{(10)}, x_8^{(10)}, x_9^{(10)})$

"how we guess"

if j ⇒ Continuous

if j ⇒ categorical

$\boxed{\text{Analogy}}$