Session 08

K-Nearest Neighbors(KNN)

Machine Learning | Zahra Amini

Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

https://zil.ink/zahraamini

Regression:

x_1	x_2	x_3	x_4	y	
–	–	–	–	–	train
–	–	–	–	–	
–	–	–	–	–	
–	–	–	–	–	test
–	–	–	–	–	

$x^{new} \rightarrow x_1 \quad x_2 \quad x_3 \quad x_4 \rightarrow y$

Regression $\longrightarrow$ Continues

Classification:

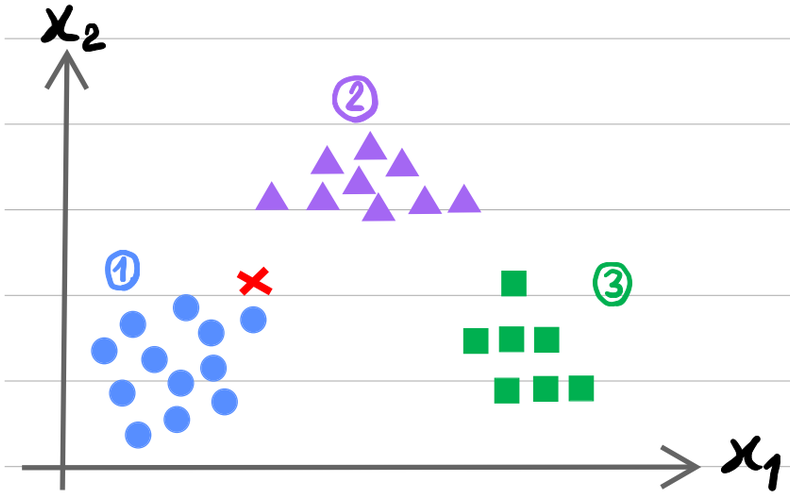
Iris
- → Iris-Setosa → Class 1
- → Iris-Versicolor → Class 2
- → Iris-Virginica → Class 3

	C1	C2	C3
y →	☐	☐	☐

Classification → Discrete

Classification
- → kNN → Supervised learning
- → Decision Tree
- → Logistic Regression
- → SVM

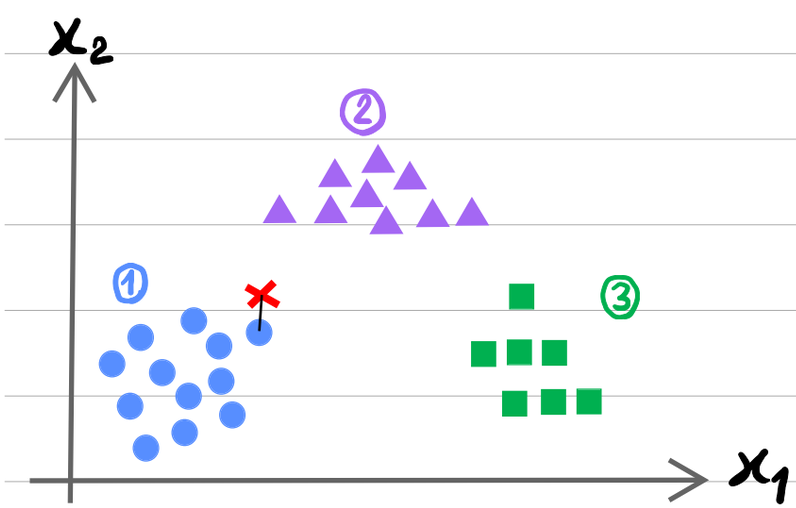
KNN (k-Nearest Neighbors)



ما می‌خواهیم پیش بینی کنیم که داده جدید (×) عضو کدام کلاس است

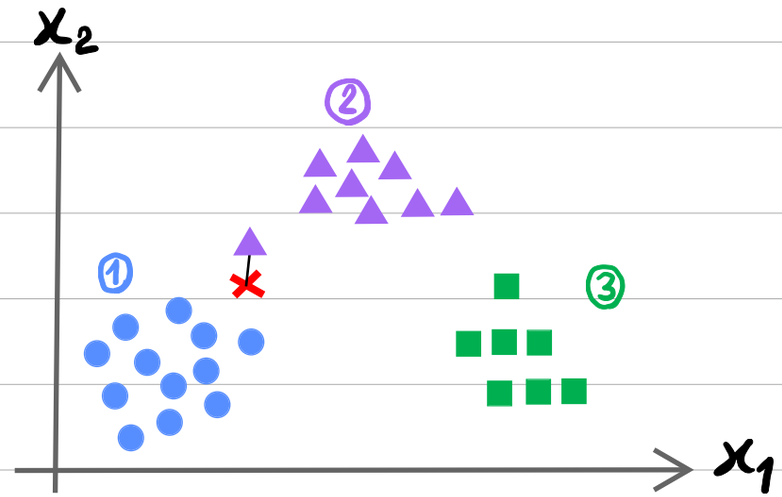
$$\begin{bmatrix} x_1^1 & x_2^1 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \\ x_1^4 & x_2^4 \end{bmatrix} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} : \begin{bmatrix} 2 \\ 2 \\ 1 \\ 3 \end{bmatrix}$$

1-NN



1. محاسبه شباهت (فاصله) داده تست با تک تک داده های train
2. پیدا کردن نزدیکترین همسایه از داده آموزش به نمونه تست
3. تصمیم گیری ← x^{new} عضو کلاسی ست که نزدیکترین همسایه اش عضو آن است.

؟: آیا با 1-NN همه چیز خوب است؟

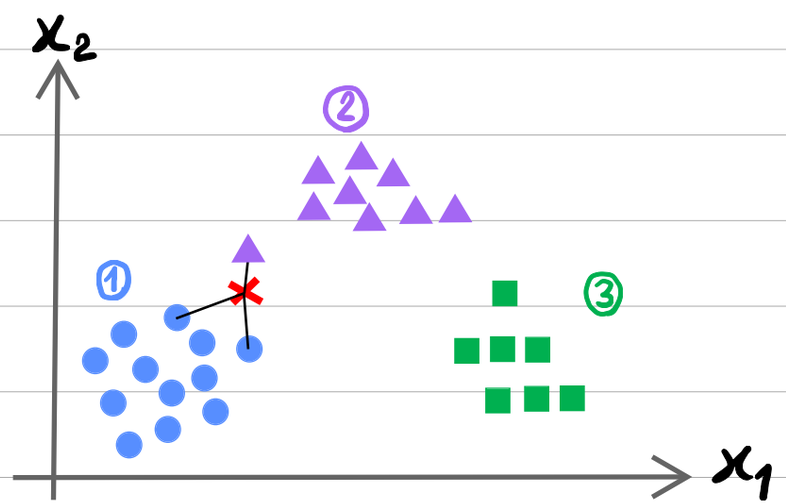


مشکل 1-NN چیست؟

معمولاً داده‌های ما زیاد تمیز نیستند و Noisy هستند.

راه حل: 1 داده‌های آموزش را کاملاً تمیز کنیم.

2 تعداد همسایه‌ها را زیاد کنیم. $(k>1)$



$k=3 \longrightarrow$	C1	C2	C3
	2	1	0

$\checkmark \longrightarrow \hat{y} = 1$

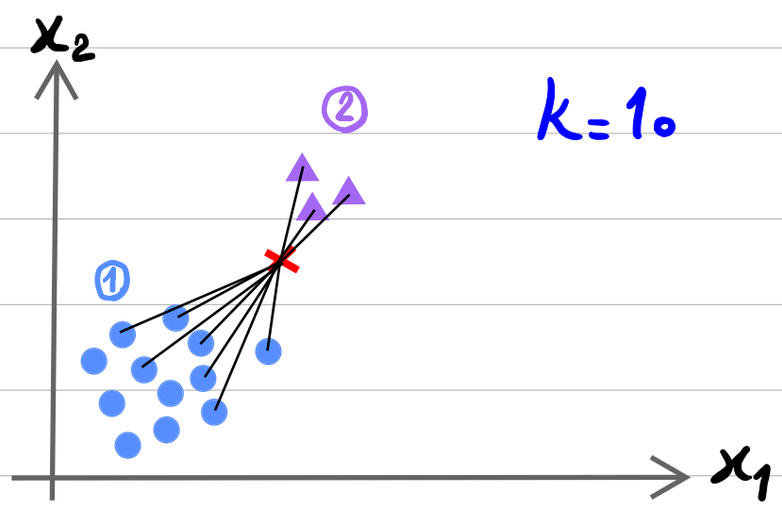
با زیاد کردن تعداد نزدیکترین همسایه ها (k) می توانیم مشکل 1-NN را حل کنیم، ولی تا چه حدی می توانیم k را زیاد کنیم؟

مشکل k بزرگ:

اگر یک کلاس تعداد نمونه train زیادی داشته باشد مثلاً 1000 تا، اما کلاس دیگر 10 تا

و با توجه به اینکه داده هایمان آنقدر تمیز نیستند که کلاس ها با هم فاصله زیاد داشته باشند ممکن است داده تست

را عضو کلاس با نمونه زیاد در نظر بگیرد در حالی که این اشتباه است.



انتخاب k بهینه بسیار مهم است. ← k ها ابرپارامتر است.

:kNN

① تعیین تعداد نزدیکترین همسایه‌ها ($k=?$)

② محاسبه‌ی فاصله‌ی نمونه تست با تمام داده‌های train

③ مرتب کردن داده‌های train براساس فاصله

④ انتخاب k تا نزدیکترین همسایه به نمونه تست

⑤ رأی‌گیری ← x عضو کلاسی است که بیشترین همسایه نسبت به نمونه‌های train را آن کلاس داشته باشد.

$k=5 \rightsquigarrow$

C1	C2	C3
3	1	1

$\rightsquigarrow \hat{y}=1 \qquad x \in C1$

؟: اگر مسأله‌مان ۲ کلاسه باشد، بهتر است k را فرد انتخاب کنیم یا زوج؟ آیا تفاوتی دارد؟

$k=4$

- C1: 2
- C2: 2

$\rightarrow \hat{y}=?$ $\quad x \in ?$

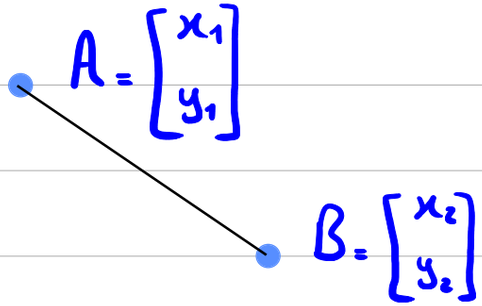
$k=5$

- C1: 5 $\rightarrow \hat{y}=1$ $\quad x \in C1$
- C2: 5 $\rightarrow \hat{y}=2$ $\quad x \in C2$
- C1: 3 C2: 2 $\rightarrow \hat{y}=1$ $\quad x \in C1$
- C1: 2 C2: 3 $\rightarrow \hat{y}=2$ $\quad x \in C2$

k فرد برای مسأله‌ای که ۲ کلاسه است بهتر است.

؟: آیا KNN فاز training دارد؟

معیار شباهت (فاصله):



① فاصله اقلیدسی Euclidean

$$dis(A,B) = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$$

فاصله ↑ شباهت ↓

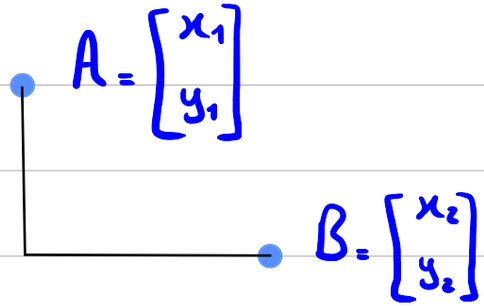
$$dis(A,B) = \sqrt{(A-B)^2}$$

فاصله ↓ شباهت ↑

$$(A-B) = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1-x_2 \\ y_1-y_2 \end{bmatrix} \longrightarrow (A-B)^2 = \begin{bmatrix} (x_1-x_2)^2 \\ (y_1-y_2)^2 \end{bmatrix}$$

$$\xrightarrow{\Sigma} \Sigma (A-B)^2 = \left[(x_1-x_2)^2 + (y_1-y_2)^2\right] \xrightarrow{\sqrt{\ }} \sqrt{\left[(x_1-x_2)^2 + (y_1-y_2)^2\right]}$$

② Manhattan / City Block بلوک شهری



$$dis(A,B) = |x_1 - x_2| + |y_1 - y_2|$$

$$dis(A,B) = \sum |A - B|$$

③ Chebychev

$$dis(A,B) = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \longrightarrow |A-B| = \begin{array}{c} r_1 \\ \begin{bmatrix} |x_1 - x_2| \\ |y_1 - y_2| \end{bmatrix} \\ r_2 \end{array}$$

$$\longrightarrow Max(r_1, r_2)$$

Minkowski ④

$$dis(A,B) = \sqrt[p]{|x_1 - x_2|^p + |y_1 - y_2|^p}$$

$$dis(A,B) = \sqrt[p]{\sum |A-B|^p}$$

- $P=1 \rightsquigarrow$ بلوک شهری
- $P=2 \rightsquigarrow$ اقلیدسی
- $P=\infty \rightsquigarrow$ chebychev

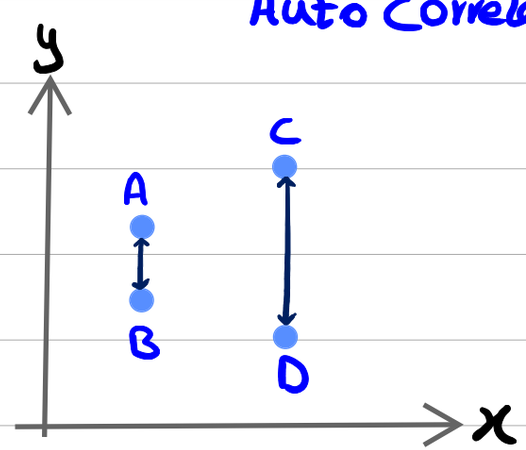
فاصله Cosine ⑤

$$dis = 1 - \frac{A'B}{\sqrt{\underbrace{(A'A)}_{\text{Auto Correlation}} * (B'B)}}$$

$A'B \rightarrow$ Cross Correlation

$$A = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}_{2\times1} \quad B = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}_{2\times1}$$

$$A'B \rightsquigarrow \begin{bmatrix} x_1 & y_1 \end{bmatrix}_{1\times2} \times \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}_{2\times1} = \begin{bmatrix} \underbrace{\quad}_{\text{Scaler}} \end{bmatrix}_{1\times1}$$



A با B و C با D باتوجه به معیار فاصله اقلیدسی متفاوت هستند اما باتوجه به معیار cosine خیر متفاوت نیستند. ← cosine اختلاف دامنه را حساب نمی‌کند و رفتار را بررسی می‌کند.

⑥ همبستگی Correlation

$$\rho(A,B) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right)$$

$$\begin{bmatrix} \quad \end{bmatrix}^T_{2\times 1} \qquad \begin{bmatrix} \quad \end{bmatrix}_{2\times 1} \longrightarrow \begin{bmatrix} \quad \end{bmatrix}$$

Scaler

در واقع همون cov است اما cov که نرمال شده است.

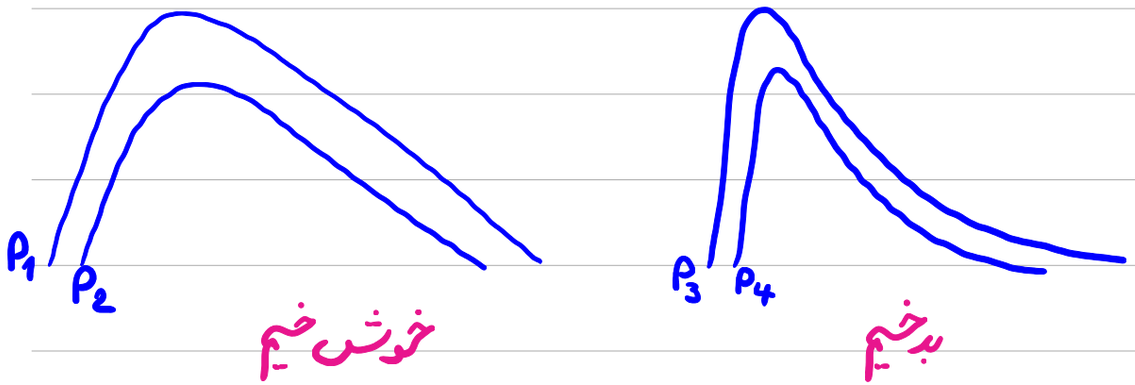
$$\rho(A,B) = \frac{Cov(A,B)}{\sigma_A \, \sigma_B}$$

ترم نرمالایز

$$Cov(A,B) = \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \mu_A) * (B_i - \mu_B)$$

$$dis(A,B) = 1 - \rho(A,B)$$

در Correlation چون تغییرات دامنه تأثیر ندارد، این معیار برای مقایسه رفتار داده‌ها کاربردی ست.



اگر معیار را فاصله‌ی اقلیدسی در نظر بگیریم

P_1 با P_2 و P_3 با P_4 در یک کلاس

قرار نمی‌گرفتند اما اگر معیار corr باشد چون نسبت به تغییرات دامنه حساس نیست بهتر عمل می‌کند.

؟: چه معیاری برای هر dataset مناسب است؟

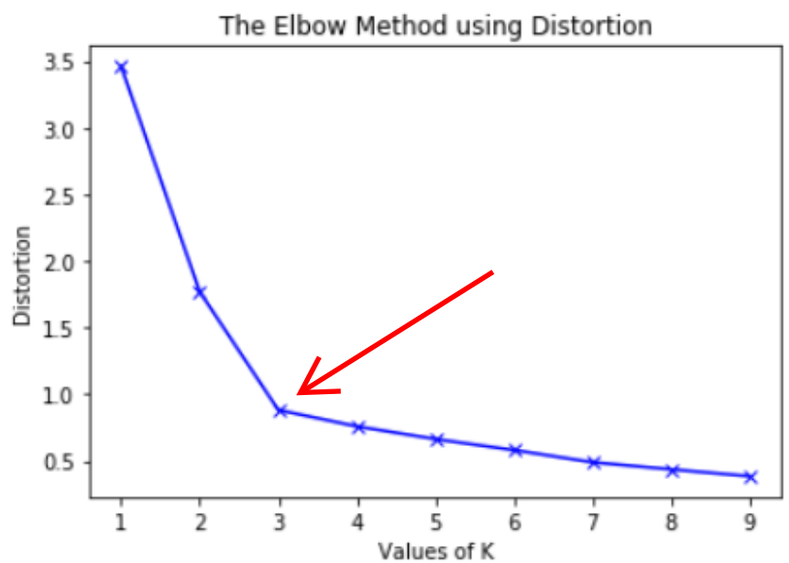
گاهی با بررسی شرایط مسأله و تحلیل شرایط می‌توانیم انتخاب صحیح داشته باشیم اما گاهی اوقات هم باید معیارهای مختلف را امتحان کنیم و هرکدام نتیجه بهتری به ما بدهد آن را انتخاب می‌کنیم.

خب در KNN دغدغه‌ی ما انتخاب مقدار k هست. چه طور k بهینه را انتخاب کنیم؟

$$\text{error rate} = \frac{1}{m}\sum_{i=1}^{m} I\left(\hat{y}^{(i)} \neq y^{(i)}\right)$$

① با استفاده از روش elbow

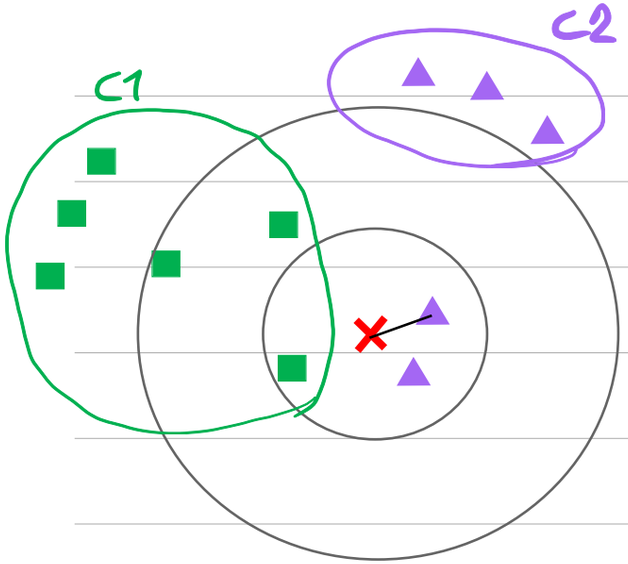
$$I\left(\hat{y}^{(i)} \neq y^{(i)}\right) \begin{cases} I(\hat{y} \neq y) = 1 \\ I(\hat{y} = y) = 0 \end{cases}$$



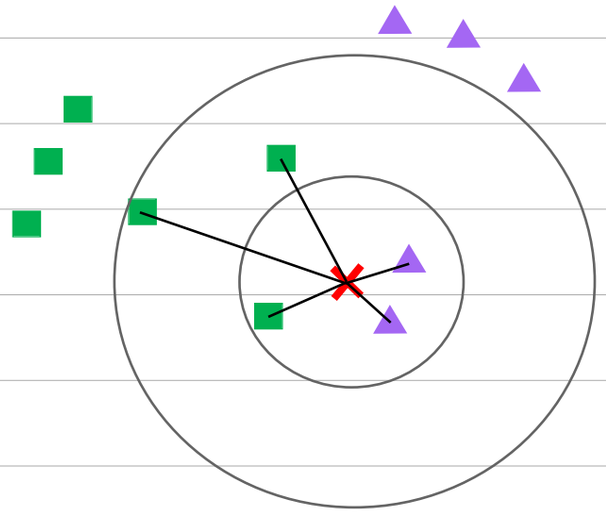
② می‌دانیم که KNN به مقدار k حساس است. ← روشی که این حساسیت را کاهش دهد می‌تواند مقدار k را هم بهینه کند

بنابراین سراغ WKNN می‌رویم.

WkNN یا kNN وزن دار :

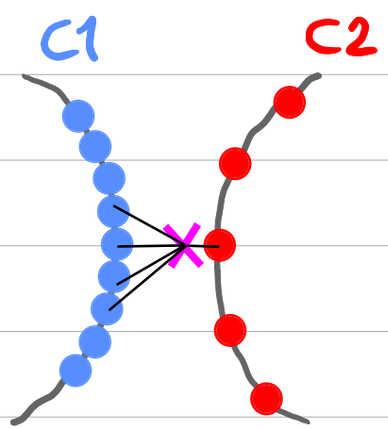


$k=1$



$k=5$

اگر به شکل کلی کلاس C2 توجه کنید می بینید که حضور داده های ▲ در مرکز شبیه نویز هست. بنابراین می توان



با افزایش k از 1 به 5 این مشکل را حل کرد.

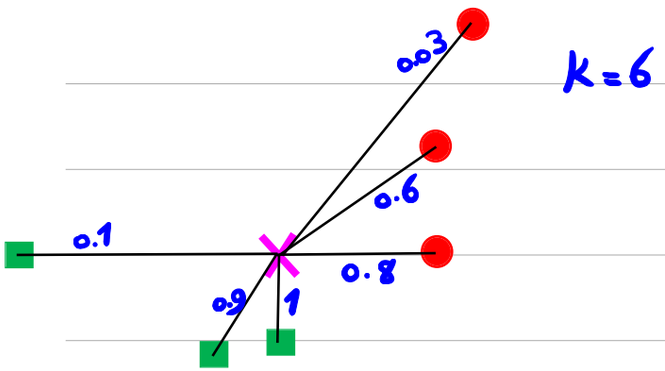
$k=5$

C1	C2
4	1

$\rightsquigarrow \hat{y}=1 \quad x \in C1 \quad X$

* چون توازن بین کلاس ها برقرار نیست بزرگ کردن k راه حل درستی نیست.

برای بهبود kNN از روشی استفاده می‌کنیم که در آن نقش نمونه‌ی آموزش نزدیکتر به نمونه تست مهم‌تر از نمونه دورتر باشد.



$k=6$

① k را عدد بزرگی در نظر می‌گیریم که نویز روی آن تأثیرگذار نباشد.

② تأثیر نمونه‌های نزدیکتر بیشتر شود. ← وزن دارش می‌کنیم

؟: چه طور kNN را وزن دار کنیم؟

فاصله‌ی نمونه test را با تک تک نزدیکترین همسایه‌ها به دست می‌آوریم. سپس dis را به Weight تبدیل می‌کنیم.

Weight میزان شباهت نمونه همسایه به نمونه test را مشخص می‌کند.

$0 \leq Weight \leq 1$

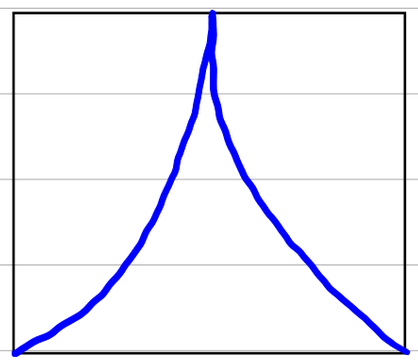
وزندار کردن:

① فاصله‌ی نمونه test را با تک تک نمونه‌های train به دست می‌آوریم.

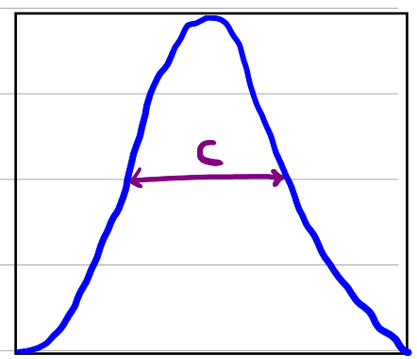
② k تا از نزدیکترین همسایه‌ها را انتخاب می‌کنیم.

③ میزان شباهت نمونه test را با تک تک همسایه‌ها بر اساس dis آنها با نمونه test محاسبه می‌کنیم.

$$W_{x_i,x_j} = \frac{1}{dis(x_i,x_j)}$$

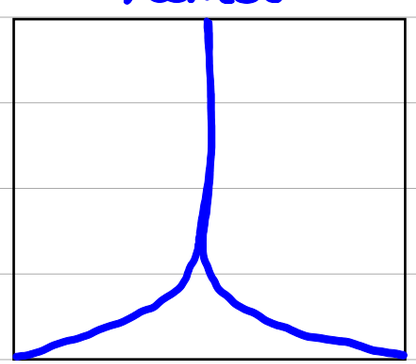


$$W_{x_i,x_j} = \frac{1}{c+(dis(x_i,x_j))^2}$$

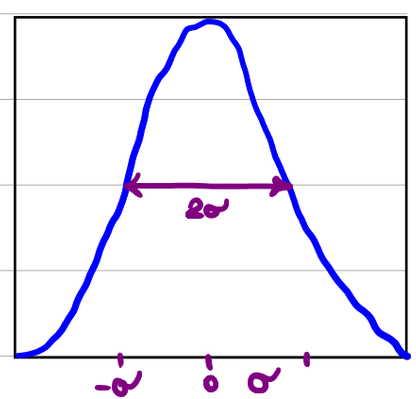


kernel

$$W_{x_i,x_j} = \frac{1}{(dis(x_i,x_j))^2}$$



$$W_{x_i,x_j} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(dis(x_i,x_j))^2}{\sigma^2}\right)$$

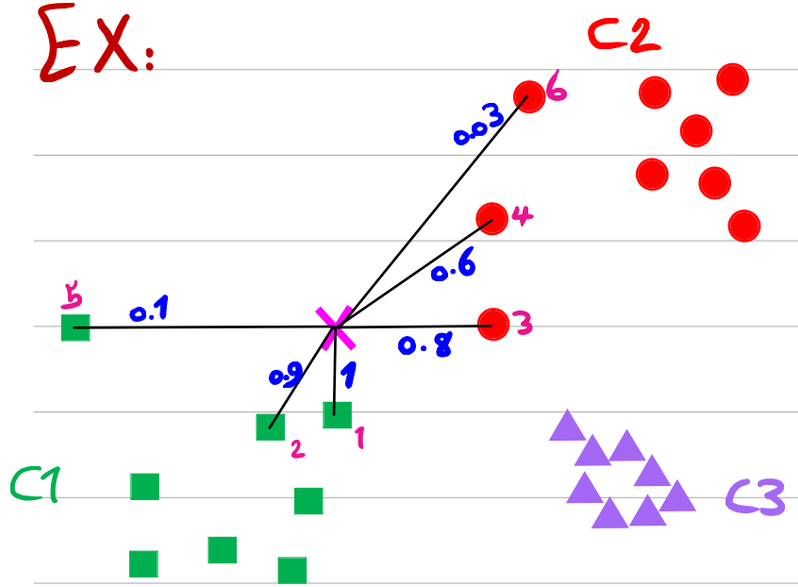


④ رأی گیری وزن دار

$$y_p(j) = \frac{\sum_{i=1}^{k} w_{x_i, x_j} * y^{(i)}}{\sum_{i=1}^{k} w_{x_i, x_j}}$$

نرمالسازی $\rightsquigarrow \frac{w_1}{\sum w}, \frac{w_2}{\sum w}, \dots, \frac{w_k}{\sum w}$

EX:



$k=6 \rightsquigarrow$ C1 C2 C3

$w_1 + w_2 + w_5$ $\quad$ $w_3 + w_4 + w_6$ $\quad$ 0

$\underbrace{1 + 0.9 + 0.1}_{=2}$ $\quad$ $\underbrace{0.8 + 0.6 + 0.03}_{1.43}$

$$\text{label } 6\text{-NN} = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \end{array} \begin{bmatrix} 1 & 1 & 2 & 2 & 1 & 2 \end{bmatrix}$$

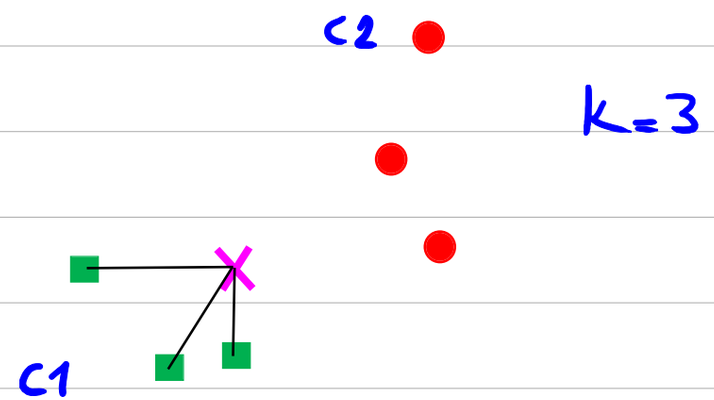
$$W = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$\xrightarrow{C1} \sum \text{label} * W$

$\cancel{w_1 \times 1}^{1} + \cancel{w_2 \times 1}^{0.9} + \cancel{w_5 \times 1}^{0.1} = 2$

؟: آیا می توان از کلاسبند KNN برای پیش بینی مقادیر پیوسته (رگرسیون) استفاده کرد؟

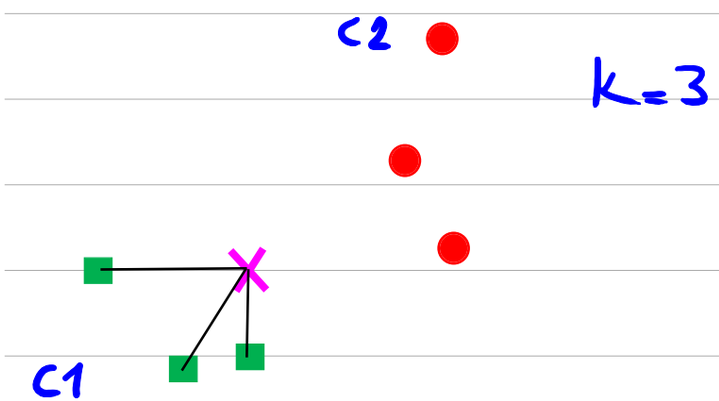
نحوه تبدیل KNN گسسته به پیوسته (تبدیل classification به Regression)



$k=3$

1	1	1

$\longrightarrow x \in C1$



$k=3$

4.5	14.36	5.25

$$\hat{y} = \frac{4.5 + 14.36 + 5.25}{3} = 8.03$$

نحوه تبدیل WkNN گسسته به پیوسته:

y های نزدیکترین همسایه‌ها	4.5	14.36	5.25	11.24	8.23
w های نزدیکترین همسایه‌ها	0.54	0.17	0.12	0.09	0.08

$$\hat{y} = (4.5 \times 0.54) + (14.36 \times 0.17) + (5.25 \times 0.12) + (11.24 \times 0.09)$$

$$+ (8.23 \times 0.08) = 7.17$$

$$\sum w = 1$$

نکته بسیار مهم، جمع wها باید 1 شود.

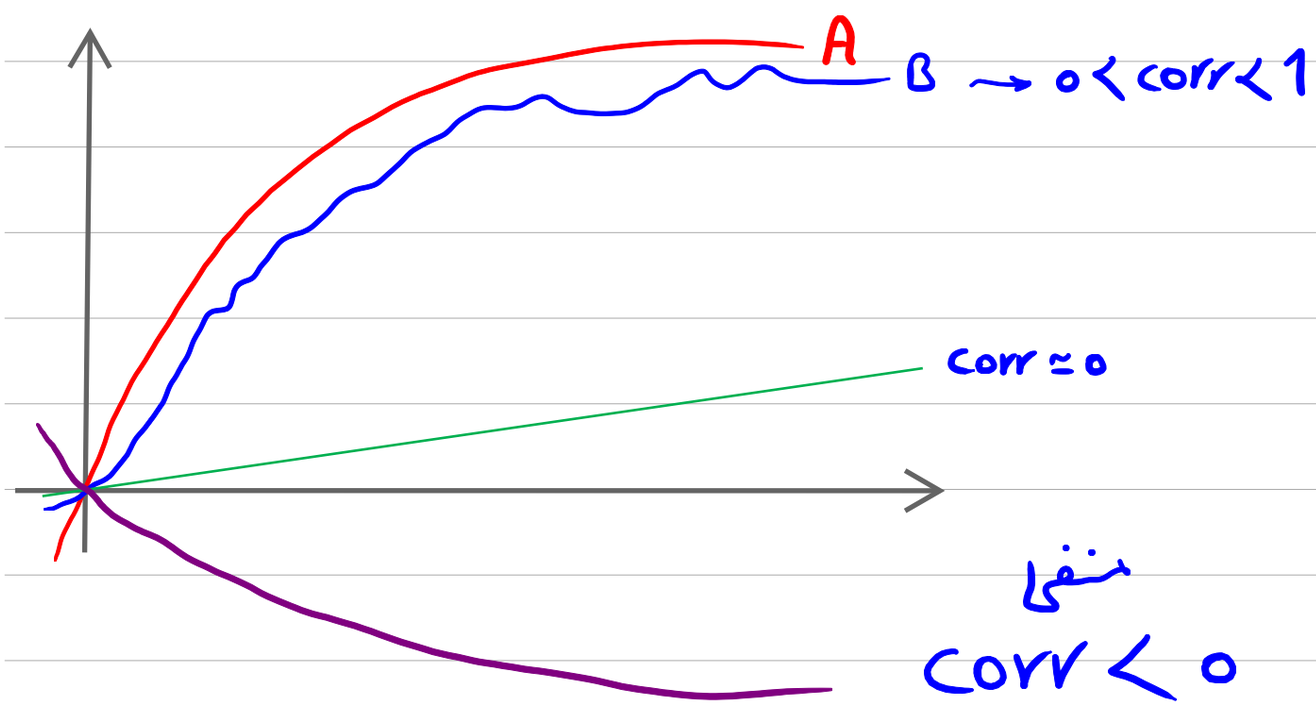
پارامتر ارزیابی KNN-Regression :

Correlation:

$$\rho(A,B) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right)$$

$$\rho(A,B) = \frac{Cov(A,B)}{\sigma_A \, \sigma_B}$$

Corr رفتار B را نسبت به A می‌سنجد



* ما قبلاً دیدیم که رنج داده‌ها بر الگوریتم‌هایی که با فاصله کار می‌کنند تأثیر می‌گذارد.

بنابراین چون اساس کار KNN با فاصله است، حتماً ابتدا داده‌هایتان را نرمال یا استاندارد کنید.