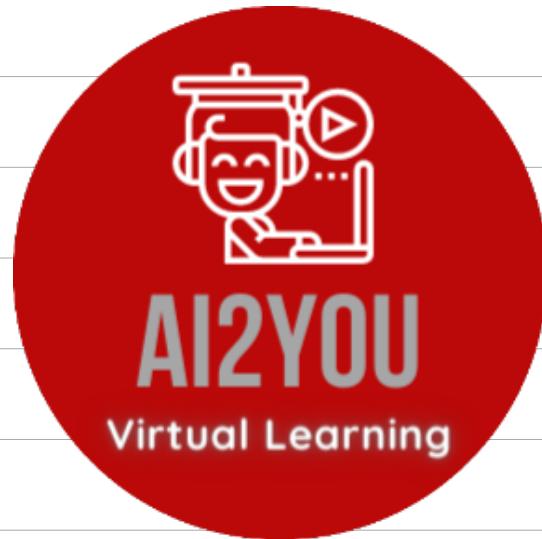


Session 7

KNN



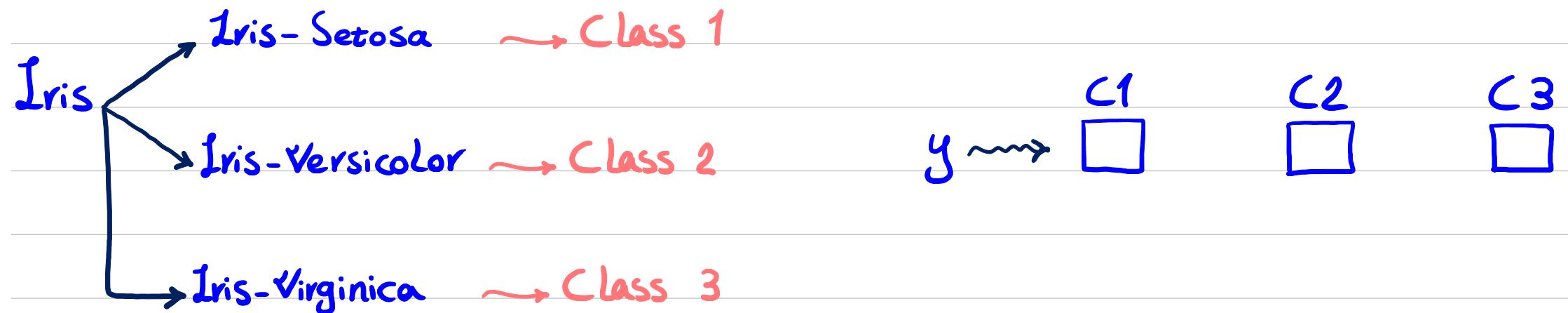
Regression:

x_1	x_2	x_3	x_4	y	
-	-	-	-	-	train
-	-	-	-	-	
-	-	-	-	-	
-	-	-	-	-	
-	-	-	-	-	

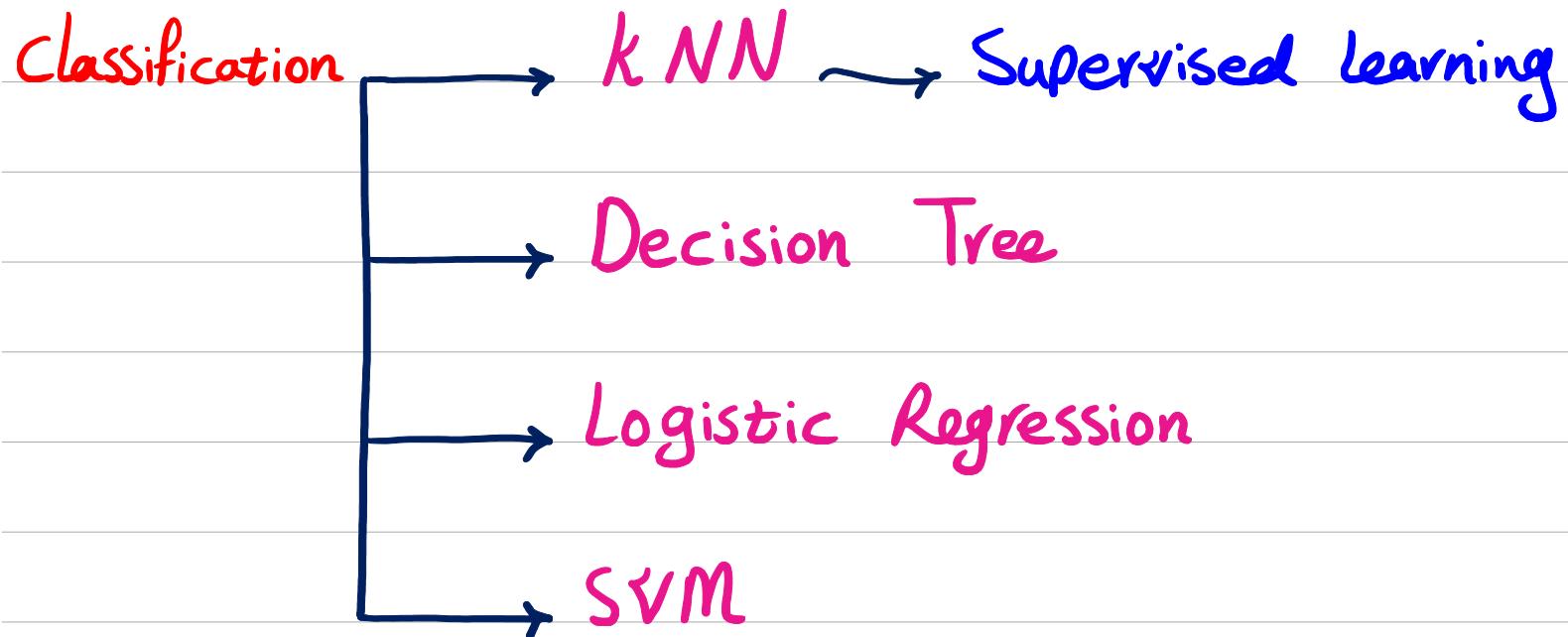
$x^{\text{new}} \rightarrow x_1 \quad x_2 \quad x_3 \quad x_4 \rightsquigarrow y$

Regression \longrightarrow Continues

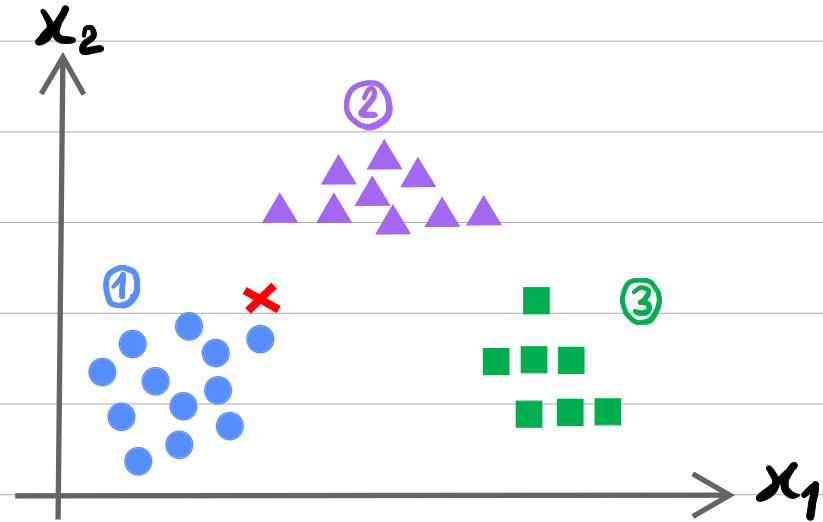
Classification:



Classification → Discrete



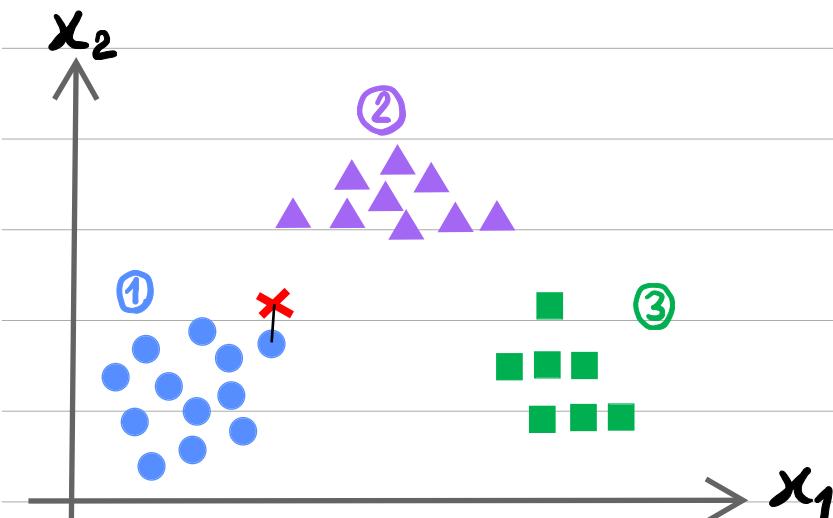
kNN (k-Nearest Neighbors)



حالی خواهیم پیش بینی کنیم که داده جدید (x) 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



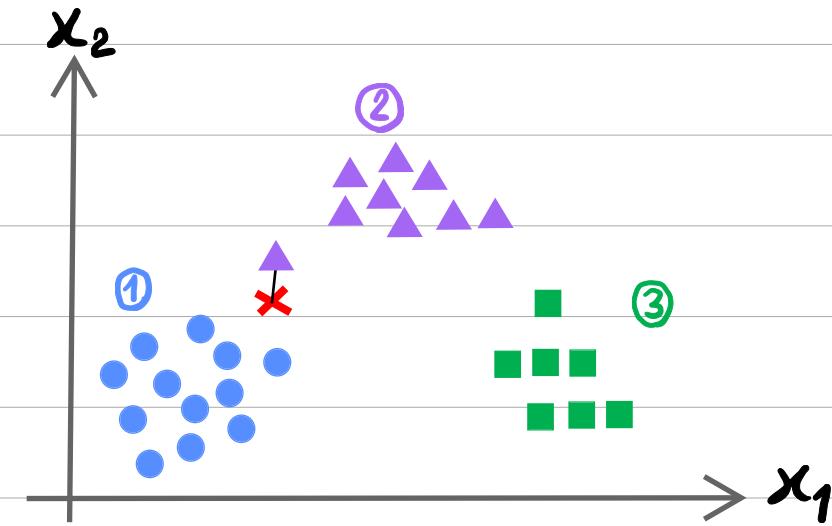
① کاربر شاخصت (فاحمله) داده تست با تک تک داده های train

② پیدا کردن نزدیکترین همسایه از داده آموزش به مخونه تست

③ تعمیم لیری \rightarrow k عضو کلاسی است که نزدیکترین

همسایه اش عضو آن است.

سؤال ۱: با ۱-NN چیز خوب است؟

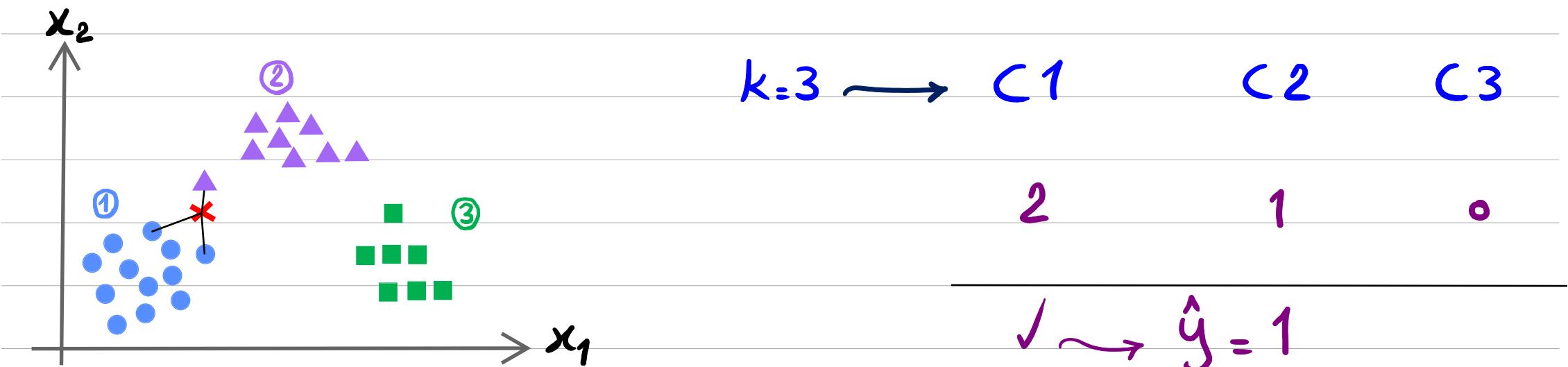


شکل ۱-NN چیست؟

مسئله داده های مازیار تکیز نموده Noisy هستند.

راه حل: ۱) داده های آموزش را کاملاً تکیز کنیم.

۲) تعداد هایی هارا مازیار کنیم. ($k > 1$)



بازیاد کردن تعداد نزدیکترین همسایه‌ها (k) می‌توانیم مشکل $NN-1$ -NN را حل کنیم، ولی تاچه حدی می‌توانیم که را زیاد کنیم؟

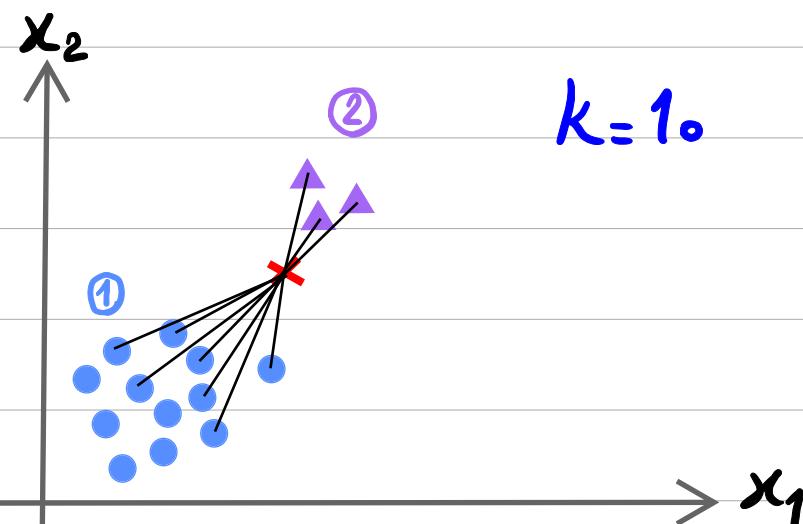
شکل k بزرگ:

اگر یک کلاس تعداد مخونه train زیادی داشته باشد مثلاً ۱۰۰۰ تا، اما کلاس دیگر ۱۰ تا

و با توجه به اینکه داده‌هایمان آنتدر تغییر نیستند که کلاس‌ها باهم فاصله زیاد داشته باشند می‌کنیم ابتدا داده‌تست

را معرفی کلاس باخونه زیاد در نظر بگیرد در حالی که این اشتباه است.

انتخاب k بحیینه سیار هم است. $\leftarrow k$ های پس پارامترات.



: kNN

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

② حسابی فاصله‌ی مخونه تست با تمام داده‌های train

③ مرتب کردن داده‌های train براساس فاصله

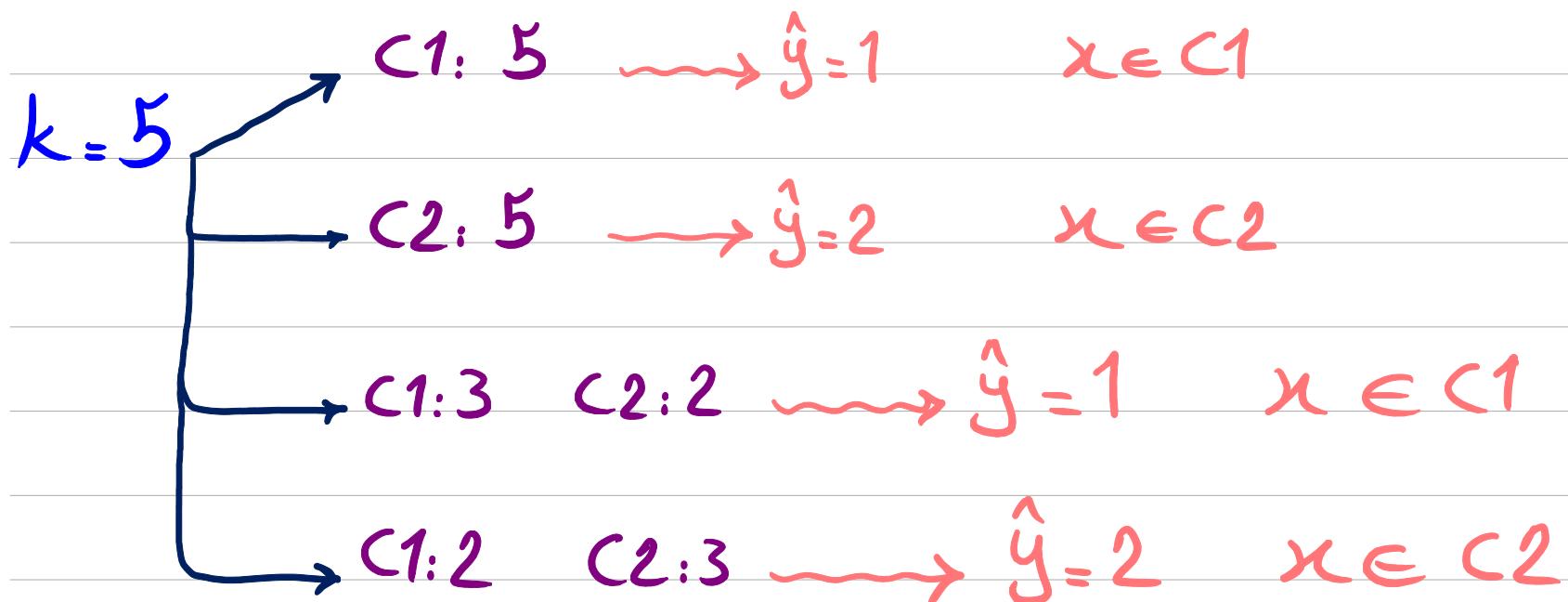
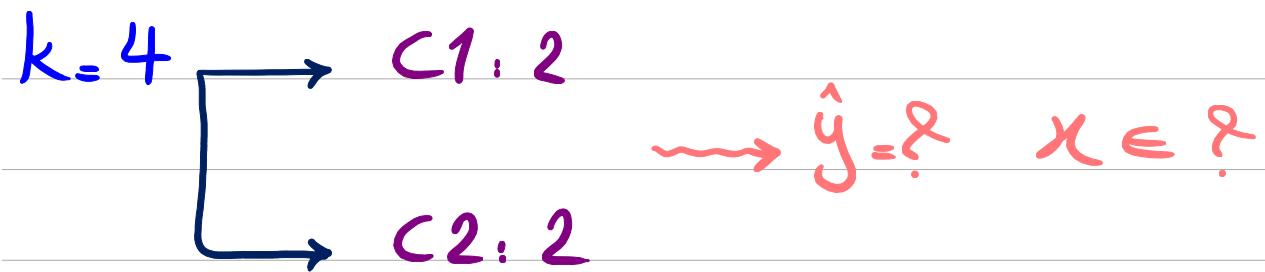
④ انتخاب k تا نزدیکترین همسایه به مخونه تست

⑤ رانی‌گیری \rightarrow در ععنو کلاسی است که بیشترین همسایه نسبت به مخونه‌های train را آن کلاس داشته باشد.

$\xrightarrow{k=5 \rightsquigarrow C_1 \quad C_2 \quad C_3 \quad \rightsquigarrow \hat{g}=1 \quad x \in C_1}$

\rightsquigarrow	$k=5 \rightsquigarrow$	C_1	C_2	C_3	$\rightsquigarrow \hat{g}=1$	$x \in C_1$
		3	1	1		

؟: اگر سواله مان یک کلاس باشد، بحث راست که افراد انتخاب کنیم یا ازدج بود آیا تفاضلی دارد؟



کند برای سواله ای که یک کلاس را بحث راست.

؟: training فاز kNN دارد؟

$$A = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$B = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

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

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

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

$$\sum (A - B)^2 = [(x_1 - x_2)^2 + (y_1 - y_2)^2] \rightarrow \sqrt{[(x_1 - x_2)^2 + (y_1 - y_2)^2]}$$

میکار شباہت (فاملہ) :

① فاملہ اقلیدسی

شباہت $\left\{ \begin{array}{l} \text{فاملہ} \\ \uparrow \end{array} \right\}$

شباہت $\left\{ \begin{array}{l} \text{فاملہ} \\ \downarrow \end{array} \right\}$

$$A = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

بلوک شهری ② Manhattan / City Block

$$B = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$\text{dis}(A, B) = |x_1 - x_2| + |y_1 - y_2|$$

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

$$\text{dis}(A, B) = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$\rightarrow |A - B| = \boxed{|x_1 - x_2|} \\ \boxed{|y_1 - y_2|}$$

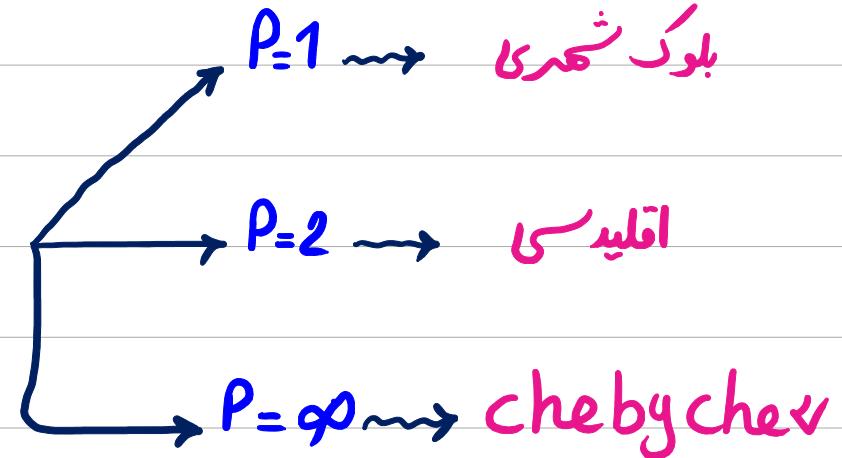
$$\hookrightarrow \text{Max}(r_1, r_2)$$

Chebychev ③

Minkowski ④

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

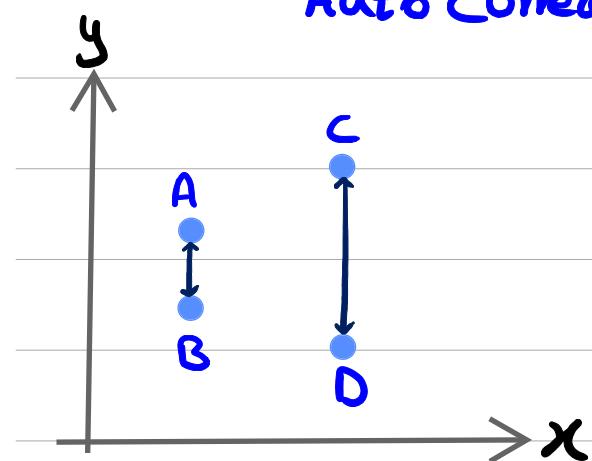
$$dis(A, B) = \sqrt[p]{\sum |A - B|^p}$$



Cosine نامه ⑤

$$dis = 1 - \frac{\sqrt{(A'A) * (B'B)}}{\sqrt{(A'A)} * \sqrt{(B'B)}}$$

Auto Correlation Cross Correlation



$$A = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}_{2 \times 1} \quad B = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}_{2 \times 1}$$

$$A'B \rightarrow \begin{bmatrix} x_1 & y_1 \end{bmatrix}_{1 \times 2} \times \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}_{2 \times 1} = \begin{bmatrix} \{ \} \end{bmatrix}_{1 \times 1}$$

Scaler

معيار cosine باتوجه به معيار فاصله اقلیدسی مسافت هسته اما باتوجه به

معيار cosine خير مسافت نسته. ← اختلاف داشته راهاب نباشد.
ورتفار رابرسي نباشد.

Correlation ⑥ مبکری

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

$$\begin{bmatrix} & \\ & \end{bmatrix}_{2 \times 1}^T \quad \begin{bmatrix} & \\ & \end{bmatrix}_{2 \times 1} \rightarrow \begin{bmatrix} & \\ & \end{bmatrix}$$

Scaler

در واقع هر دو cov انتسابی cov نهاده شده است.

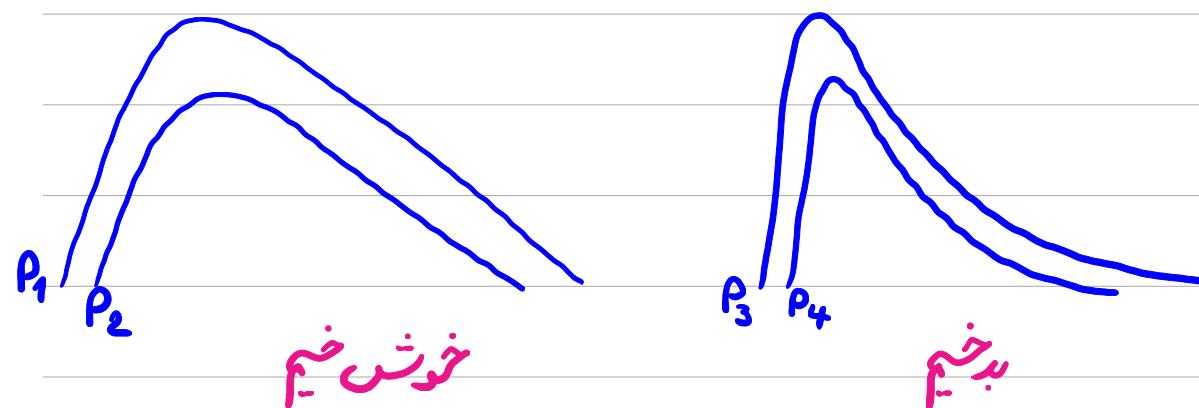
$$\rho(A, B) = \frac{\text{cov}(A, B)}{\sigma_A \sigma_B}$$

Term نرم نرم الاینر

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

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

در Correlation چون تغییرات داشته تأثیر ندارد، این معیار برای مقایسه رفتارداده‌ها



کاربردی است.

اگر معیار را ناصله‌ی اقلیدسی در نظر بگیریم

P_4 با P_3 و P_2 با P_1 در یک کلاس

قرار نبی از فتنه اما اگر معیار corr باشد چون نسبت به تغییرات داشته حساس نیست بکثر عمل می‌کند.

؟: چه معیاری برای dataset مناسب است؟

گاهی با بررسی شرایط مسئله و تحلیل شرایطی آنرا نم انتخاب صحیح داشته باشیم اما گاهی اوقات هم

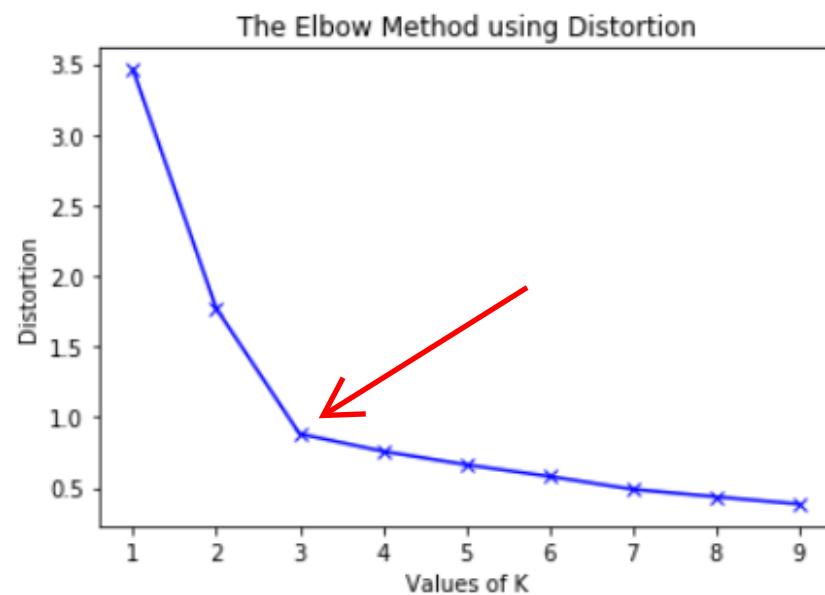
باید معیارهای مختلف را استخان کنیم و هر کدام نتیجه بکتری به حابدهد آن را انتخاب می‌کنیم.

خب در kNN دغدغه‌ی می‌انتساب مقدار k است. چه طور k بهینه‌رالنتیاب کنیم؟

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

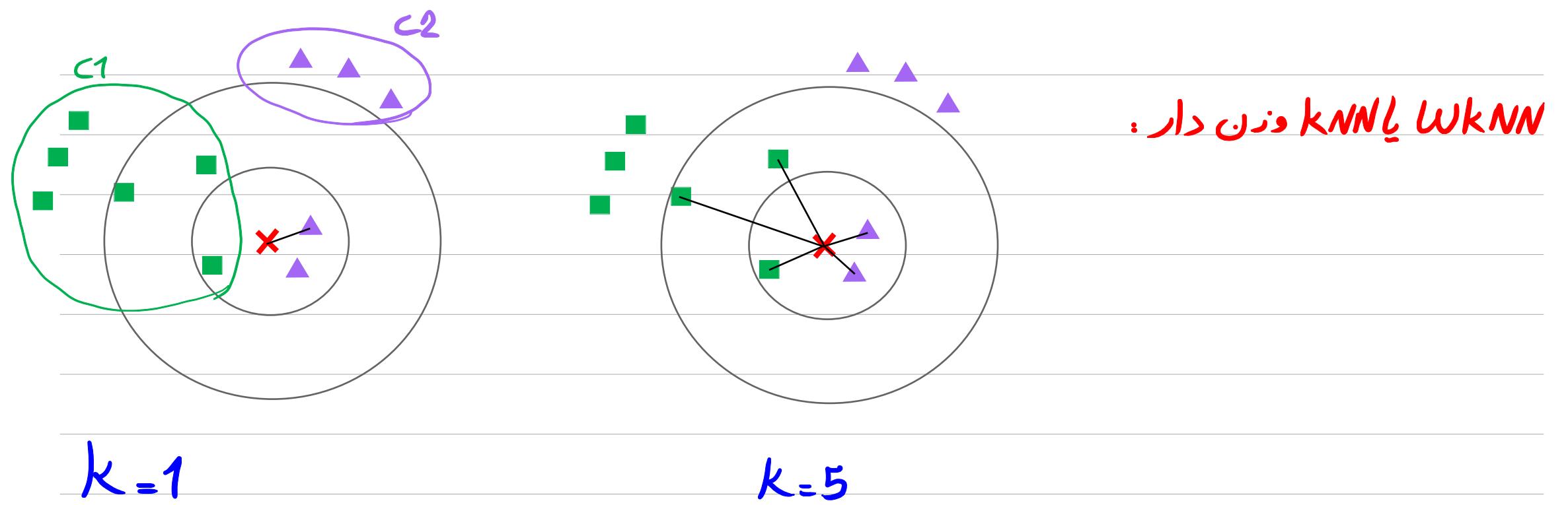
۱) استفاده از روش elbow

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



۲) دانست که مقدار k را با محاسبات را کاملاً داده‌ی تواند مقدار k را بهینه کنند.

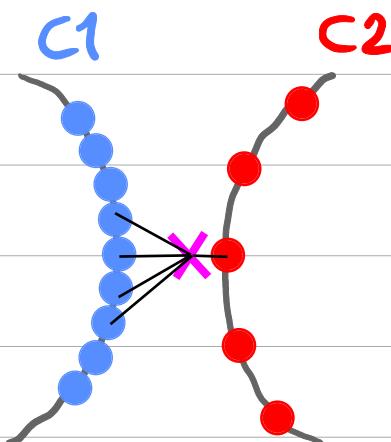
بنابراین سلسله $WkNN$ را درویم.



kNN و $WkNN$ دار:

آخر ب شکل که کلاس ۲ توجه کنیم که حضور داده های ▲ در مرکز شبیه نویز هست. بنابراین ممکن است

با افزایش k از ۱ به ۵ این مشکل را حل کرد.



$k=5$

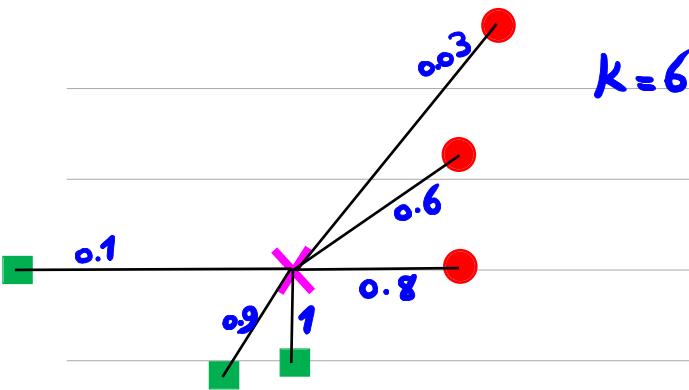
C_1
4

C_2
1

$\hat{y}=1 \quad x \in C_1 \quad X$

* چون توانان بین کلاس ها برقرار نیت بزرگ کردن k راه حل درستی است.

برای بحث kNN از روشهای استناده‌ی کنیم که در آن نقش نمونه‌ی آموزش نزدیکتر به نمونه است هم‌تر از نمونه دورتر باشد.



① k را عدد بزرگی در نظر مایلیم که نویز روی آن تاثیرگذار نباشد.

② تاثیر نمونه‌های نزدیک‌تر بر شرودر دهن دارش می‌کنیم

؟ چه طور kNN را دهن دار کنیم ؟

فاصله‌ی نمونه test را با تک نزدیک‌ترین همایه‌ها به دست می‌آوریم. سپس dis را به weight تبدیل می‌کنیم.

$$0 \leqslant \text{Weight} \leqslant 1$$

میزان شباهت نمونه همایه به نمونه test را مشخص می‌کند.

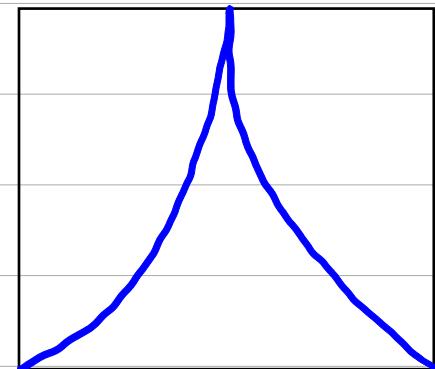
وزندار لردن:

① فاصله‌ی x_{test} را با تک مخونه‌های $train$ به دست می‌آوریم.

② k -تا از نزدیکترین های $train$ را انتخاب می‌کنیم.

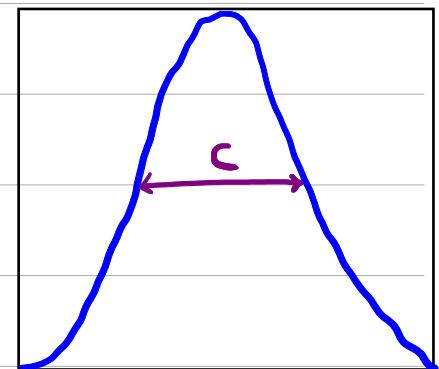
③ میزان شباهت مخونه x_{test} را با تک های $train$ با مخونه x_{test} محاسبه می‌کنیم.

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

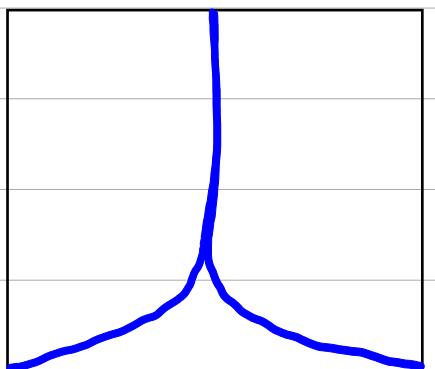


kernel

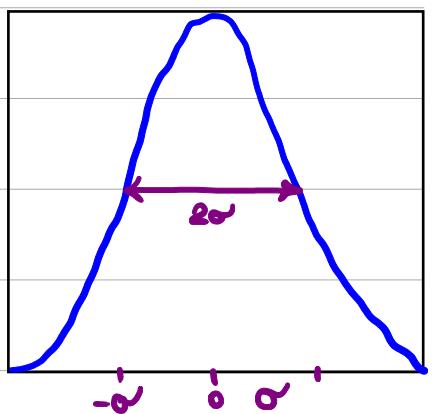
$$W_{x_i, x_j} = \frac{1}{C + (\text{dis}(x_i, x_j))^2}$$



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



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

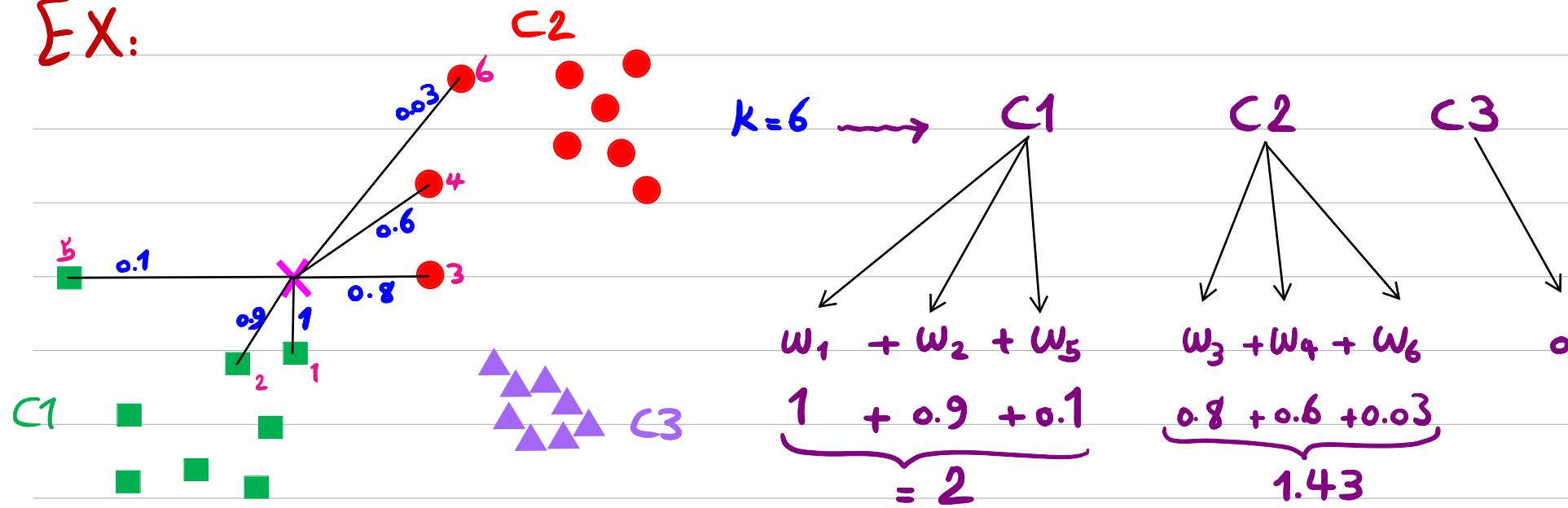


رای اکثری فوزن دار ④

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

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

EX:



$$\text{label 6-NN} = [1 \ 1 \ 2 \ 2 \ 1 \ 2]$$

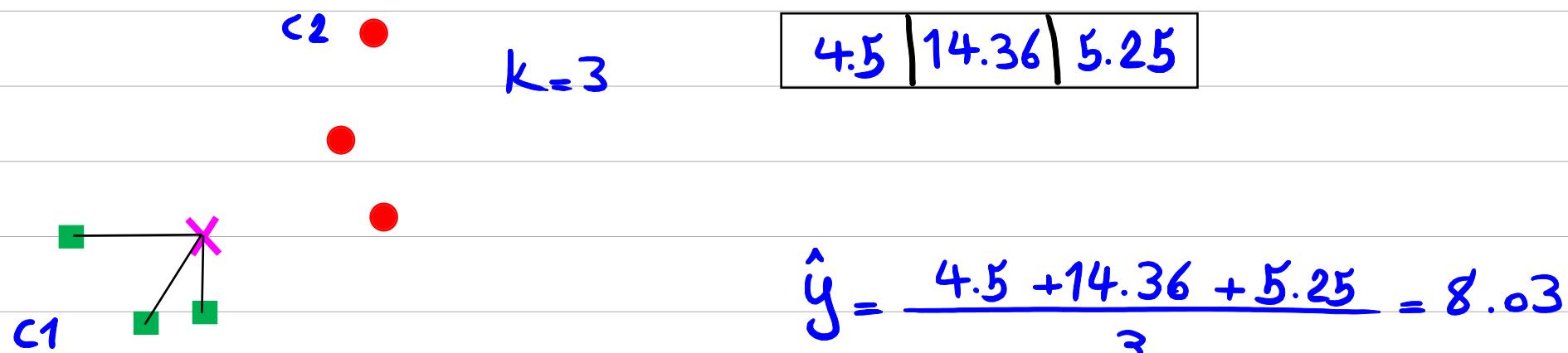
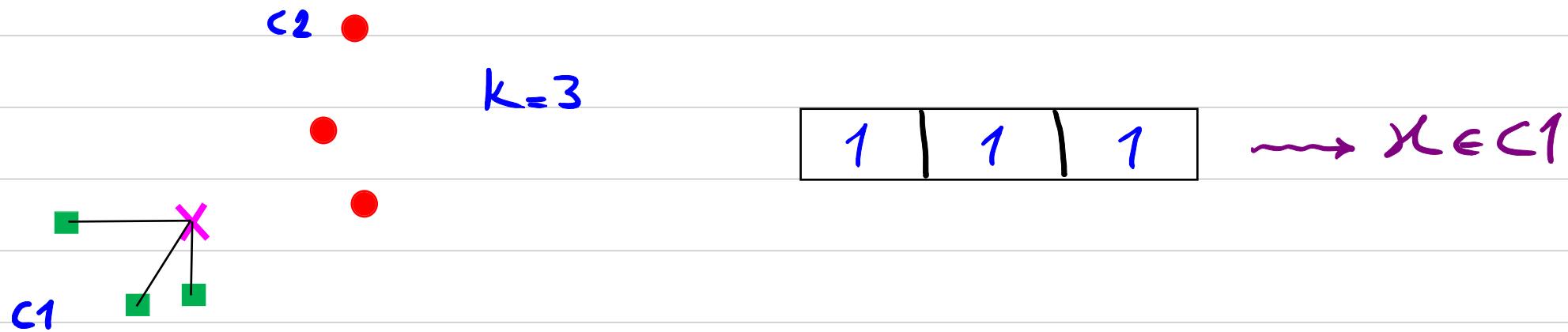
$$W = [1 \ 1 \ 0 \ 0 \ 1 \ 0]$$

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

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

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

نحوه تبدیل kNN کسته به پیوسته (تبدیل Regression ~ classification)



نحوه تبدیل W_{kNN} کرته ب پرسه:

و های نزدیکترین مسایعها

4.5	14.36	5.25	11.24	8.23
-----	-------	------	-------	------

W های نزدیکترین مسایعها

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 هم جمع w ها با 1 شود.

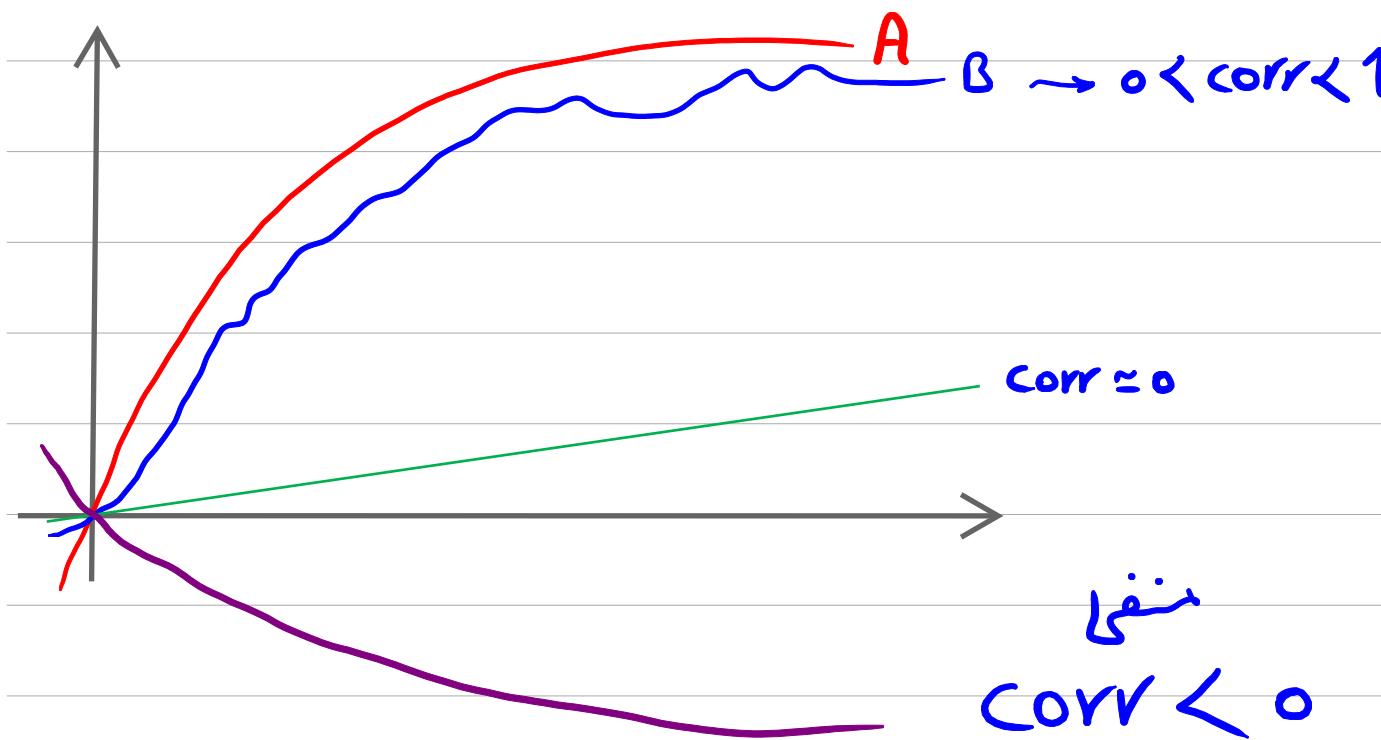
پارامتر ارزیابی : KNN-Regression

Correlation:

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

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

نکته: اگر A و B را مترابه دانستیم، آنگاه $\rho(A, B) = 1$



* ما قبلاً دیدیم که رنج داده‌ها بر الگوریتم‌هایی که با فاصله کاری کنند تاثیر عی لذارد.

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

