

AC-Resum.pdf



user_2397943



Aprenentatge Computacional



3º Grado en Ingeniería de Datos



**Escuela de Ingeniería
Universidad Autónoma de Barcelona**



Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

→ Plan Turbo: barato

→ Planes pro: más coins

pierdo espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



AC

2023-24

Anàlisi computacional

WUOLAH

WUOLAH

ÍNDIX DE CONTINGUTS

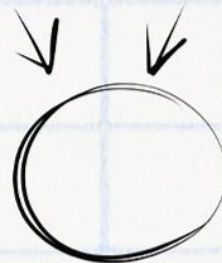
Conceptes previs	9
Aprenentatge Supervisat vs. No supervisat	9
k-NN	10
Diagrama de Voronoi	11
Classificació	11
Escarlar	12
Normalització	12
Inconvenients del model kNN	12
Introducció al model d'avaluació.....	13
Groundtruth	13
Matriu de confusió	13
Model de Decisió.....	15
Millor tall de threshold.....	16
Representació de mesures	17
Procés experimental	20
Error de substitució	20

Imagínate aprobando el examen

Necesitas tiempo y concentración

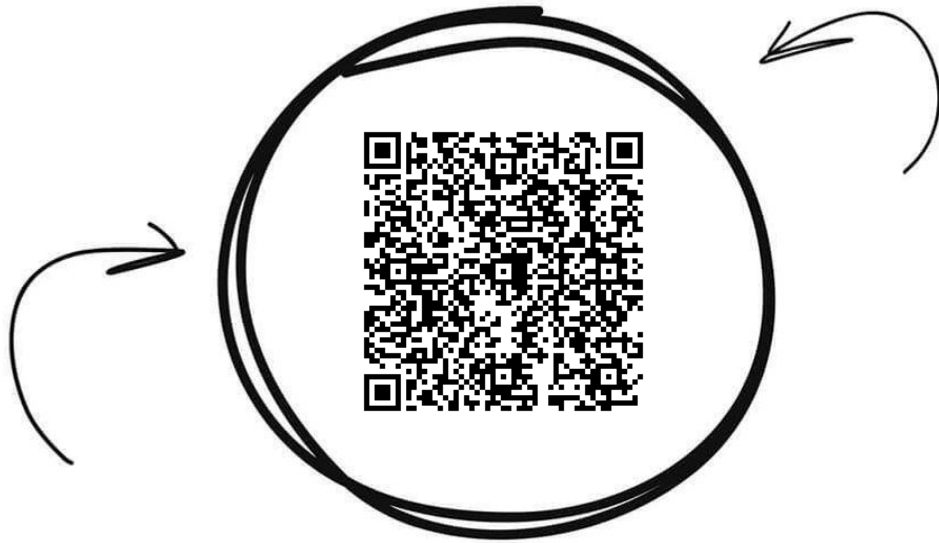
Planes	 PLAN TURBO	 PLAN PRO	 PLAN PRO+
 Descargas sin publi al mes	10 	40 	80 
 Elimina el video entre descargas			
 Descarga carpetas			
 Descarga archivos grandes			
 Visualiza apuntes online sin publi			
 Elimina toda la publi web			
 Precios Anual <input type="checkbox"/>	0,99 € / mes	3,99 € / mes	7,99 € / mes

Ahora que puedes conseguirlo, ¿Qué nota vas a sacar?



WUOLAH

Aprenentatge Computacional



Banco de apuntes de la

WUOLAH



Comparte estos flyers en tu clase y consigue más dinero y recompensas

- 1** Imprime esta hoja
- 2** Recorta por la mitad
- 3** Coloca en un lugar visible para que tus compis puedan escanar y acceder a apuntes
- 4** Llévate dinero por cada descarga de los documentos descargados a través de tu QR



Biaix VS Variància	21
Validació creuada	23
deixar un a fora	23
Validació d'arrencada (bootstrapping)	23
Regressió I	25
Regressió lineal	25
Funció de pèrdua.....	26
Descens del gradient	27
Derivades de la funció d'error quadràtic	28
Funció de cost	29
Altres algoritmes	30
Múltiple regressió	30
Regressió polinòmica	30
Normalització de característiques	30
Equació normal	31
Regressió II	33
Regressió logística	33

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

→ Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

Decisió del límit	34
Funció de cost	35
Classificació multi classe.....	37
One VS All.....	37
One VS One	37
Models erronis	38
Underfitting.....	38
Overfitting	38
Biaix.....	39
Variància	39
Instància del problema	42
Regularització.....	44
Magnitud del coeficient	44
Regressió de Ridge	45
Funció de pèrdua Lasso - l_1	46
SVM.....	47
Outliers	48

WUOLAH

Kernel trick.....	48
Objectiu.....	49
Funcions de cost.....	50
Representació dual	51
Arbres de decisió.....	52
Construcció dels arbres	53
Mètode recursiu	53
Criteris de parada	53
Heurístiques	53
Guany d'informació	54
Índex de Gini	56
Comparativa criteris de divisió	57
Problemes	57
Poda	58
Reduced-Error Pruning	58
Pruning Pesimistic	58
Cost-complexity Pruning.....	59

Conjunts de mètodes	60
Ensembles homogenis.....	60
Mètode per canviar dades.....	60
Bagging and boosting	63
Bagging.....	63
Boosting	64
Random forest	66
Algorisme	66
Clustering.....	68
Tipus de clusters.....	68
Clustering partitional.....	69
Clustering jeràrquic	72
Validació del mètode cluster	74
Mesures internes	74
Bag of Words.....	75
<i>Dense sampling</i>	76
Representació de l'estímul	76

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

Classificació	77
SVM multiclasse	77
Kernel	78
Tipus d'assignacions	79
Variables latents	79
Sistemes recomanadors	80
Mesura de la rellevància	80
Obtenció d'informació	80
Precisió	81
Recall	81
F-measure	82
realització ranking	82
Precisió mitjana	83
Sistema recomanador	84
Filtratge col·laboratiu	84
Factorització de matrius	87
SVD	87

WUOLAH

Probabilitats	88
Probabilitat conjunta.....	88
Probabilitat condicionada	88
Regla de Bayes	89
Joint distribution	90
Naives Bayes	91
Suposició Naïve	91

CONCEPTES PREVIS

DADES

Un conjunt de registres de dades (exemples, instàncies o casos) descrit per:

- k atributs
- classes: cada exemple està etiquetat amb una classe pre-definida.

OBJECTIU

Aprendre un model classificador a partir de les dades que es poden utilitzar per predir les classes de les noves (futures o de test) instàncies o casos.

APRENTATGE SUPERVISAT VS. NO SUPERVISAT

Supervisat

La classificació es fa a partir d'exemples.

- Supervisió: Les dades són etiquetades amb classes ja definides.

No supervisat

Les etiquetes són desconegudes. I donades unes dades, la tasca és establir l'existència de classes o clústers en les dades.

Treballarem amb models supervisats.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

K-NN

Dades d'entrenament

Calcular distància

Dades de test

- Triar les k dades més properes amb una mètrica.

- A partir de les distàncies calculades a entrenament i identificar els k veïns.

A l'hora de classificar un punt (pertanyent en un conjunt de dades amb valors continus), tenim en compte els k veïns més a prop i calculem la mitjana:

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i$$

On y és la classe del punt i.

TEMPS D'ENTRENAMENT

És el temps que no requereix modificar les dades.

TEMPS D'INFERÈNCIA (test)

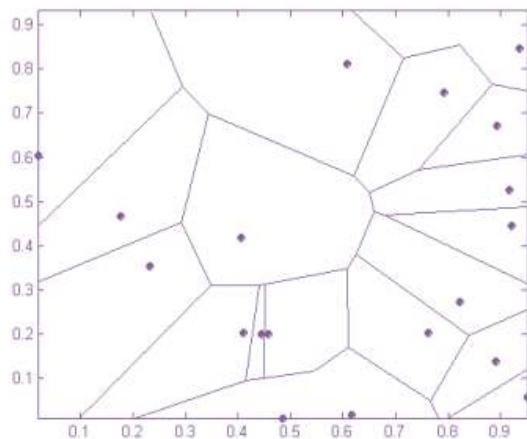
És el tems més elevat perquè s'han de fer els càlculs i és el temps per decidir quina categoria correspon cada dada.

Un model que ha après prèviament d'un conjunt dades d'entrenament, ajuda que el sistema resolgui la tasca T millor que si no hagués après prèviament. → MESURA D'EFICIÈNCIA D'UN MODEL CLASSIFICADOR

WUOLAH

DIAGRAMA DE VORONOI

Les àrees del diagrama representen la influència de cada punt.



CLASSIFICACIÓ

Per calcular la distància entre dos punts es pot fer servir l'Euclidiana:

$$d(p, q) = \sqrt{\sum_i (p_i - q_i)^2}$$

I per determinar la classe de la llista dels veïns més propers es pot atribuir un factor de pes:

$$w = \frac{1}{d^2}$$

També s'agafa la majoria de classes entre els kNN.

SELECCIÓ DE K

- Si la k és massa petita, la classificació tindrà sensibilitat amb punts sorollosos.
- Si la k és massa gran, el veïnatge (de k veïns) pot incloure punts d'altres classes.

Escalar

Els atributs es poden escalar per prevenir que algunes distàncies estiguin dominades per un atribut en concret. Ja sigui perquè treballem amb diferents unitats de mesures.

Normalització

Quan treballem amb moltes dimensions, la distància euclidiana pot ser ineficient, de manera que podem normalitzar els vectors a una unitat llargada.

A més, per resoldre possibles atributs irrellevants i per reduir la dimensionalitat del problema, podem eliminar aquests atributs no prou importants per l'estudi.

INCONVENIENTS DEL MODEL KNN

- No construeixen models explícitament.
- Classificar dades desconegudes té un cost alt.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

INTRODUCCIÓ AL MODEL D'AVUACIÓ

GROUNDTRUTH

Es pot referir al procés de comprar el píxel d'un satèl·lit és comparat amb el que és en realitat, amb la fi de verificar els continguts del píxel de la imatge.

MATRIU DE CONFUSIÓ

TP	FN	Sensitivity
FP	TN	Specificity Fall-out
Precisió		Accuracy

MESURES

Exactitud: Mesura la proporció total de X que el model ha classificat correctament.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

Accuracy = 100% → Classificació perfecta.

WUOLAH

Precisió: Mesura la proporció/qualitat dels casos detectats com a positius i que realment ho són.

$$Precision = \frac{TP}{TP + FP}$$

Precision = 100% → Tota la classificació està formada per positius. Ignorem quants positius no hem detectat.

Sensibilitat: Mesura la capacitat/eficiència del model per identificar correctament tots els casos positius. Recall = True Positive Rate

$$Recall = \frac{TP}{TP + FN}$$

Recall = 100% → Tots els positius s'han classificat com a positiu. Ignorem com els Negatius s'han classificat.

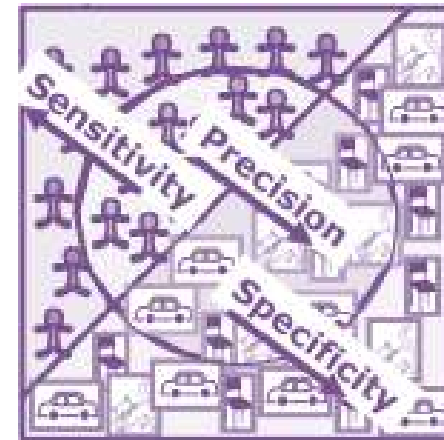
Especificitat: Eficiència en deixar a fora els casos FP

$$Fall - out = \frac{TN}{TN + FP}$$

Specificity = 100% → Tots els negatius s'han classificat com a negatius. Ignorem com els Positius s'han classificat.

On **Fall out** = False Positive Rate (rate incloent FP)

Fall out = 100% → Tots els negatius estan classificats com a Positius.



Error: Mesura la proporció total que el model ha classificat erròniament.

$$Error = \frac{(FP + FN)}{TP + FN + FP + TN}$$

F-measure:

$$F = \frac{2 * sensitivity * precision}{sens + prec}$$

MODEL DE DECISIÓ



Un model de decisió sol necessitar un *threshold* per arribar a una decisió final. Per cada valor de *thr* obtenim una matriu de confusió diferent, ja que cada valor es prenen diferents dades i el nombre de positius i negatius varia.

Un bon *threshold* és quan: $RECALL > FALL-OUT$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

MILLOR TALL DE THRESHOLD

YOUDEN INDEX

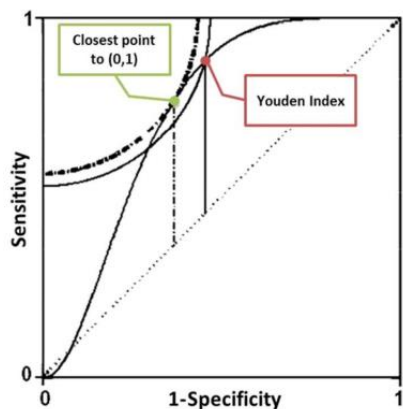
$$J = \text{sensitivity} + \text{specificity} - 1$$

$$\text{Thr} = \arg \max (J)$$

(0,1) DISTANCE

$$D = (1 - \text{sensitivity})^2 + (1 - \text{specificity})^2$$

$$\text{Thr} = \arg \min (D)$$



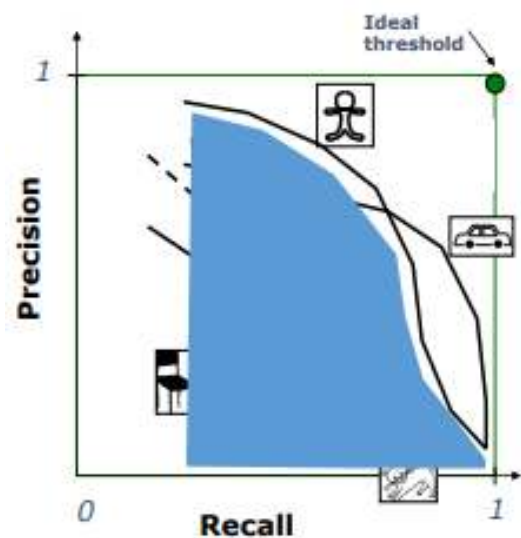
WUOLAH

REPRESENTACIÓ DE MESURES

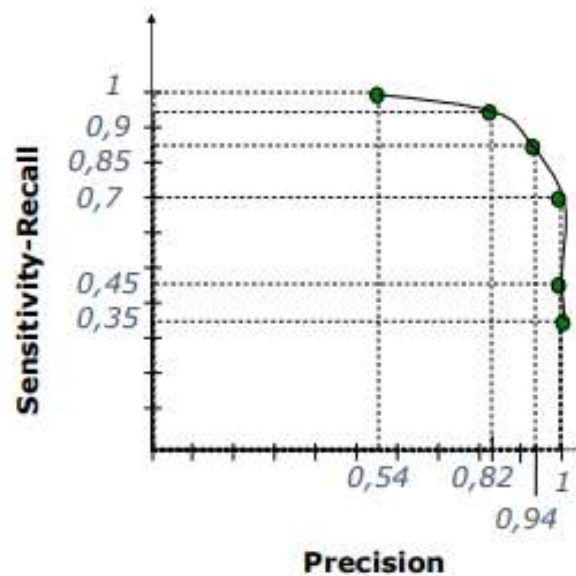
MITJANA DE PRECISIÓ

Mitjanes de precisió sobre tot el rang de recall.

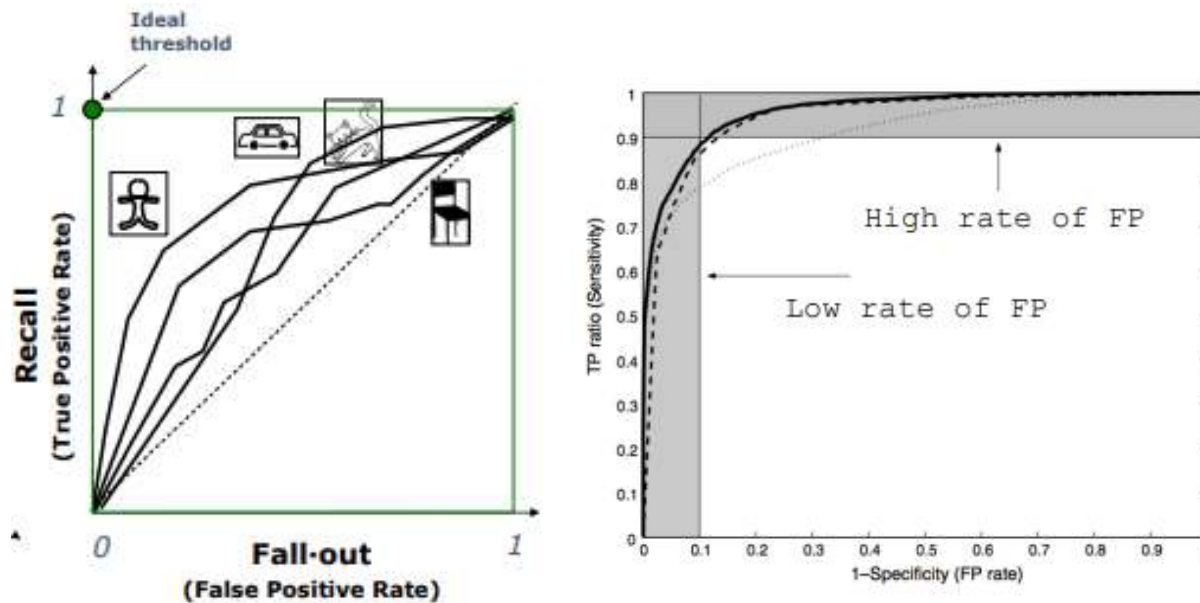
Són una bona relació per veure el pes dels FP.



SENSIBILITAT – PRECISIÓ



ROC: Receiver Operating Characteristic



L'àrea sota de la corba (AUC) és una mètrica pel perfil del classificador com a un tot, i s'utilitza per fer *ranking*. Una àrea de 1 és una classificació perfecte, 0.5 aleatòria i 0 és sinònim de tenir un classificador totalment nul.

Una classificació totalment aleatòria dona un punt al llarg de la línia diagonal, que s'anomena també línia de no-discriminació, des de l'extrem inferior esquerre fins a la cantonada superior dreta (independentment dels tipus de base positives i negatives).

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato
Planes pro: más coins

pierdo espacio

AC

2023-24

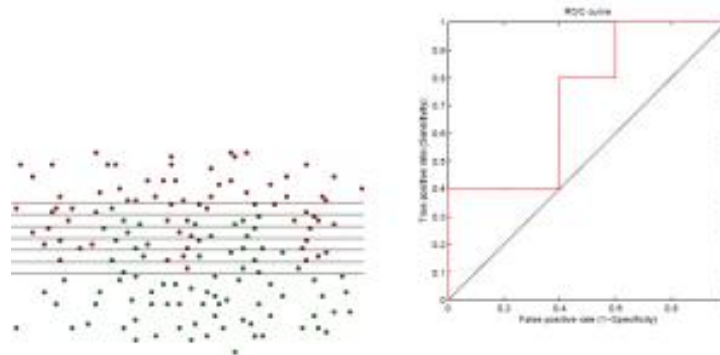
FER EXERCICI A LLIBRETA

Exercise 1: Answer the following question.

16. (JW) A system which was designed to retrieve sailing-images gave the following retrieval result:



The whole data set contains five sail boats and five non-sailboats. Draw the ROC curve (horizontal: false positive rate and vertical true positive rate).



Contingency Matrices:

N: Retrieved Im.		N = 1		N = 2	
True Pos	False Pos	1	0	2	0
False Neg	True Neg	4	5	3	5
(Recall, Fall-out)		(0.2, 0)		(0.4, 0)	
N = 3		N = 4		N = 5	
True Pos	False Pos	2	2	3	2
False Neg	True Neg	3	3	2	3
(Recall, Fall-out)		(0.4, 0.2)		(0.6, 0.4)	
N = 6		N = 7		N = 8	
True Pos	False Pos	4	3	5	3
False Neg	True Neg	1	2	0	2
(Recall, Fall-out)		(0.8, 0.6)		(1, 0.8)	
N = 9		N = 10		N = 11	
True Pos	False Pos	5	4	5	5
False Neg	True Neg	0	1	0	0
(Recall, Fall-out)		(1, 1)		(1, 1)	

WUOLAH

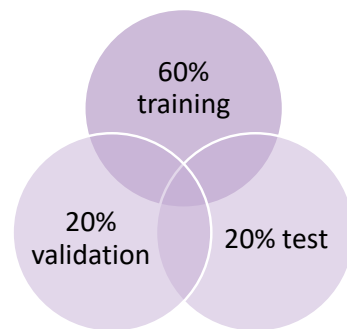
PROCÉS EXPERIMENTAL

Esperem que el classificador entrenat es pugui generalitzar i tingui una bona precisió a l'hora de fer el test.

Els conjunts d'entrenament i de test han de ser disjunts, de tal manera que cap dada pertanyi als dos.

I el conjunt de test proporciona el rendiment final del sistema. Però, si es vol mesurar l'eficiència del nostre conjunt d'entrenament i triar el model més idoni es pot utilitzar el conjunt de **validació**.

Normalment,



Els tres conjunts poden trobar-se al mateix domini o que el test o la validació estigui dins d'un altre domini o el test i la validació estiguin en un altre domini.

ERROR DE SUBSTITUCIÓ

Mesura l'error de classificació en el conjunt d'entrenament. Quan utilitzem aquest error ens podem trobar:

- El classificador entrenat pot o no pot ser prou representatiu per tot la població.
- El classificador pot aprendre particularitats que no es troben a la població: **over fitting**.

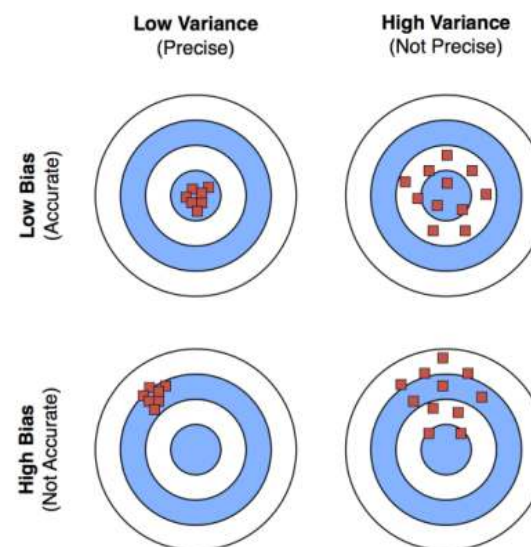
BIAIX VS VARIÀNCIA

Biaix

És un error que ve de suposicions errònies en l'aprenentatge. Alt biaix pot fer que ignorem relacions rellevants entre característiques i sortides objectiu.

Variància

És un error originari de la sensibilitat a petites fluctuacions en el conjunt d'entrenament. Alta variància comporta **overfitting**: modelar dades soroll.



Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio

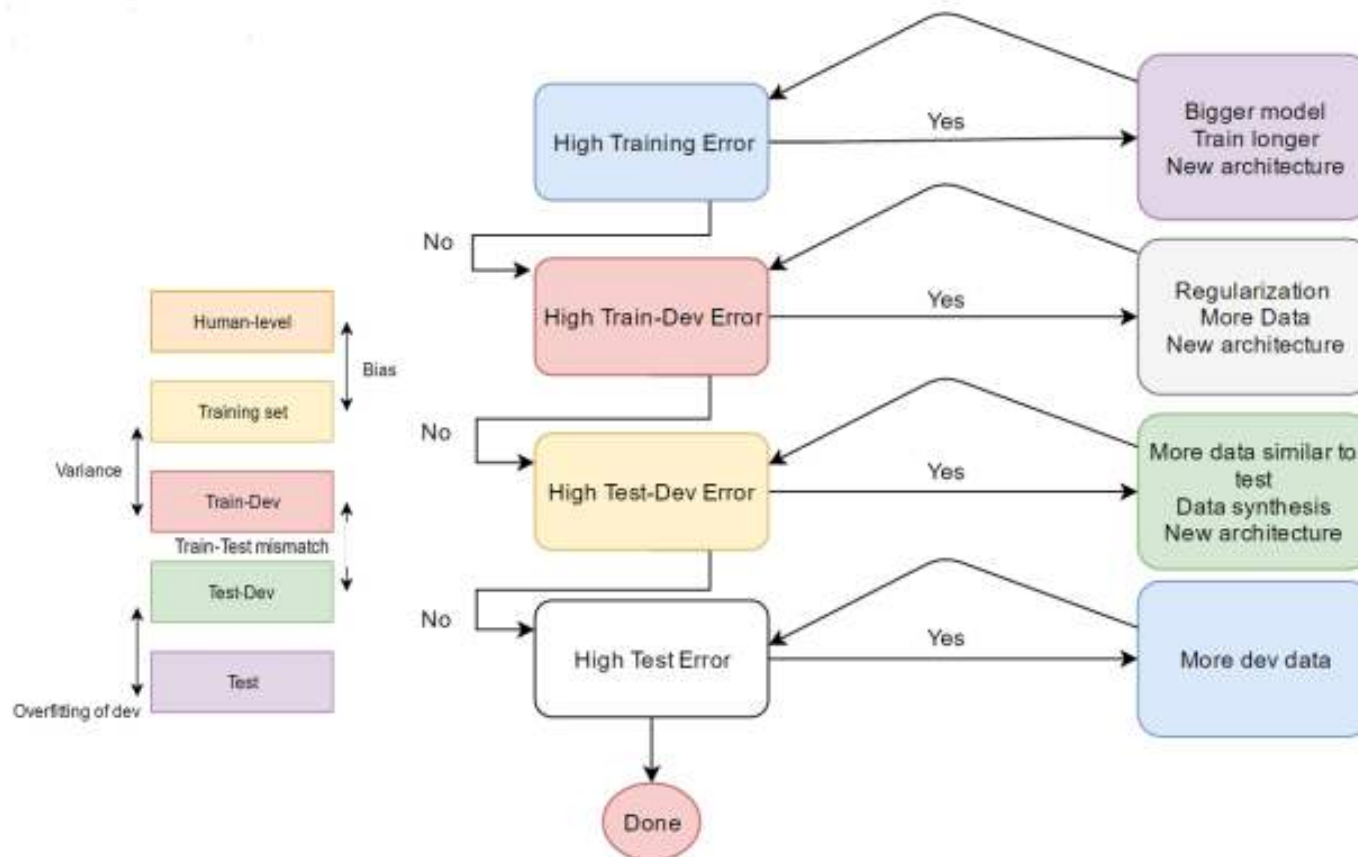


Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24



WUOLAH

VALIDACIÓ CREUADA

Separar les dades en entrenament, validació i test redueix el nombre de dades. De manera que pot afectar la generalització del classificador.

Cross validation:

1. Separar les dades aleatòriament en k subconjunts amb mida igual o similar.
2. Triar un subconjunt i entrenar el classificador amb els $k-1$ subconjunts restants.
3. Repetir k vegades fins que tots els subconjunts s'han utilitzat per testear.

El rendiment final és el resultat de la mitjana de les desviacions estàndard. Aquesta mitjana és un estimador del rendiment asimptòtic del classificador. El rendiment verdader es troba en dues distàncies de desv. estàndard amb la mitjana.

DEIXAR UN A FORA

Leave-one-out és un sistema que es tria una mostra i aquesta servirà pel test. Llavors la resta són per entrenar.

Amb aquest sistema, entren N classificadors per N mostres en el nostre conjunt d'entrenament, fet que pot representar un alt cost.

VALIDACIÓ D'ARRENCADA (BOOTSTRAPPING)

Mostreig uniforme amb substitució dels elements disponibles: un cop una mostra és seleccionada, torna a la mostra d'entrenament (es pot tornar a seleccionar). Les dades no seleccionades seran el test.

Exemple: 0.632 bootstrap = 63.2% serà conjunt d'entrenament i 36.8%, de test.

- $\left(1 - \frac{1}{x}\right)^x \approx e^{-1} = 0.368$
- $(1-1/x)x \approx e^{-1} = 0.368$

I repetint el procés k vegades:

$$acc(M) = \sum_{i=1}^k (0.632 * acc(M_i)_{test_{set}} + 0.368 * acc(M_i)_{train_{set}})$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo
espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

REGRESSIÓ I

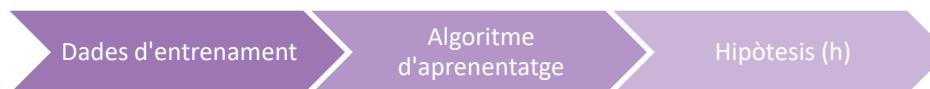
REGRESSIÓ LINEAL

La regressió és capaç de predir valors continus.

NOTACIÓ

- **m**: nombre d'exemple d'entrenament
- **x**: variable d'entrada (atribut), es pot tractar d'un vector d'informació
- **y**: variable de sortida i d'objectiu
- **(x⁽ⁱ⁾, y⁽ⁱ⁾)**: i exemple d'entrenament

PROCÉS



- A partir d'uns atributs (x) s'estima un atribut (y).

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

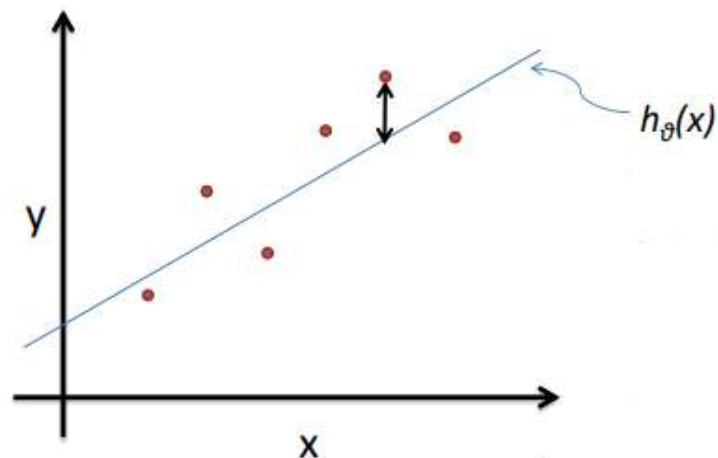
on les θ són els coeficients de cada atribut (de vegades es denota com a pes w).

Tenim varies x , amb una pendent cada una i representant un tipus de dada diferent.

WUOLAH

Funció de pèrdua

Per veure com els exemple estan de lluny de la predicció, podem fer servir l'error quadràtic (*squared error function*). De manera que la millor recta és aquella que minimitza tot els errors dels punts (x, y) .



$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2$$

Correspon a la funció de cost (*loss function*). Normalitzem amb la mitjana $(1/m)$. El cost se centra en buscar les thetes, llavors el cost computacional és molt baix.

On n és el nombre de variables:

$n=1$

$n \geq 1$

Hypothesis:

$$h_g(x) = g_0 + g_1 x$$

$$h_g(x) = g_0 x_0 + g_1 x_1 + \dots + g_n x_n$$

Parameters:

$$g_0, g_1$$

$$g_0, g_1, \dots, g_n$$

Cost Function:

$$J(g_0, g_1) = \frac{1}{2m} \sum_{i=1}^m (h_g(x^{(i)}) - y^{(i)})^2 \quad J(g_0, g_1, \dots, g_n) = \frac{1}{2m} \sum_{i=1}^m (h_g(x^{(i)}) - y^{(i)})^2$$

Goal:

$$\underset{g_0, g_1}{\text{minimise}} (J(g_0, g_1))$$

$$\underset{g_0, g_1, \dots, g_n}{\text{minimise}} (J(g_0, g_1, \dots, g_n))$$

DESCENS DEL GRADIENT

Donada una funció de cost $J(\theta_0, \theta_1)$ hem de buscar els paràmetres θ_0, θ_1 que minimitzin la funció J . Per fer-ho:

1. Es comença amb θ_0, θ_1 inicials.
2. Calcular $J(\theta_0, \theta_1)$.
3. Canviar θ_0, θ_1 per tal de reduir la funció de cost.
4. Continuar fins arribar a un mínim (local o global).

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

"Repetir fins convergència"

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \text{ amb } j = 0 \text{ i } j = 1$$

L'actualització s'ha de fer simultàniament per totes les j .

α = learning rate \rightarrow com de pressa volem arribar al mínim:

- Si és massa petita, es necessiten molts passos per arribar a la convergència.
- Si és massa gran, pot fer que la convergència vagi lenta o no s'arribi mai, pel gran canvi de punts.

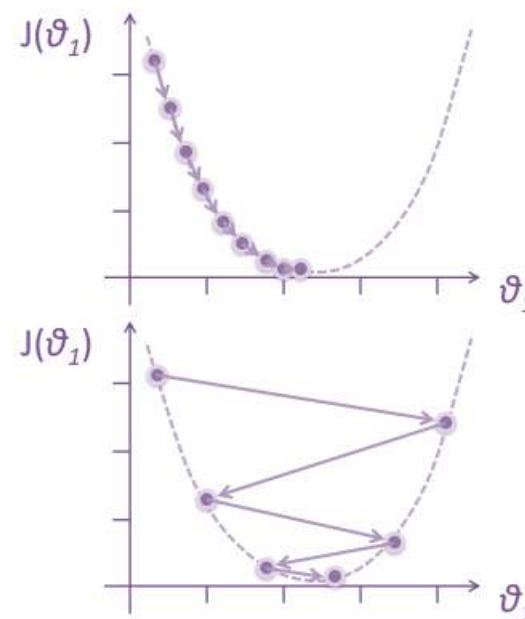
La resta de la funció és la quantitat de canvi i la direcció que segueix:

- Si la derivada és positiva vol dir que s'avança cap un direcció enrere.
- Si la derivada és negativa el pendent també ho és i per tant, s'avança cap endavant.

Derivades de la funció d'error quadràtic

La funció d'error quadràtic és convexa, per això té un sol mínim global.

S'ha de calcular la derivada i multiplicar-la per totes les thetes.



WUOLAH

$$\begin{aligned}\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) &= \frac{\partial}{\partial \vartheta_j} \frac{1}{2m} \sum_{i=1}^m (h_{\vartheta}(x^i) - y^i)^2 = \\ &= \frac{\partial}{\partial \vartheta_j} \frac{1 * 2}{2m} \sum_{i=1}^m (\vartheta_0 + \vartheta_1 x^i - y^i)^2, \text{ escribim 2 per eliminar}\end{aligned}$$

Si:

- Fem derivada de ϑ_0 : multipliquem $\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) * 1$
- Fem derivada de ϑ_1 : multipliquem $\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) * x^i$, on $x^i = \vartheta$

Funció de cost

$$\begin{aligned}J(\theta) &= \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^i), y^i) = \\ &= -\frac{1}{m} \sum_{i=1}^m [y^i \log h_{\theta}(x^i) + (1 - y^i) \log(1 - h_{\theta}(x^i))]\end{aligned}$$

I per minimitzar, com hem explicat abans, DERIVAR (amb el learning rate definit).

ALTRES ALGORITMES

Batch Gradient Descent

A cada pas del descens de gradient, utilitza tots exemples de l'entrenament, és a dir m exemples.

Stochastic Gradient Descent

A cada pas utilitza un exemple d'entrenament.

Newton's method, Conjugate gradient method, Normal equation...

MÚLTIPLE REGRESSIÓ

Aquí treballem amb més d'una característica, llavors tenim més d'una θ .

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T X, \text{ on } X \text{ matriu}$$

El terme independent (x_0) sempre val **1** (columna amb ple d'uns) i serveix perquè la funció quedi **homogènia**.

En el cas la múltiple regressió, no estem ajustant una recta, sinó que un pla.

REGRESSIÓ POLINÒMICA

Quan es treballa amb dades amb diferents unitats, amb la regressió polinòmica podem crear variables (columnes) resultants d'un producte, suma, resta de variables "existents".

NORMALITZACIÓ DE CARACTERÍSTIQUES

Fent la normalització es vol que cada característica estigui entre -1 i 1, per exemple.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

NORMALITZACIÓ DE MITJANA

Centrar les dades: cada dada li restem la mitjana de cada variable (columna en qüestió) de train i dividir-li std de cada variable (cada columna).

I sobre les dades de test també traiem la mitjana i la desviació calculades amb les de **train** (no la de test).

No s'acostuma a utilitzar MinMax perquè té en compte les dades outlier.

EQUACIÓ NORMAL

$$\theta = (X^T X)^{-1} X^T y$$

Si $X^T X$ no és invertible:

- Hi ha redundància de característiques (lineal independent).
- Hi ha masses característiques,

Llavors, **eliminem** algunes *features* o utilitzem **regularització**.

WUOLAH

DESCENS DEL GRADIENTS VS EQUACIÓ NORMAL

Quan utilitzem cada mètode:

Descens del gradient	Equació normal
Necessita triar α .	No s'ha de triar α .
Necessita moltes iteracions.	No s'ha d'iterar.
Va bé encara que n sigui gran.	$(X^T X)^{-1}$ ha de ser invertible.
Necessita normalització de característiques.	És lent si la n és gran $\rightarrow O(n^3)$
Molts mínims locals.	

REGRESSIÓ II

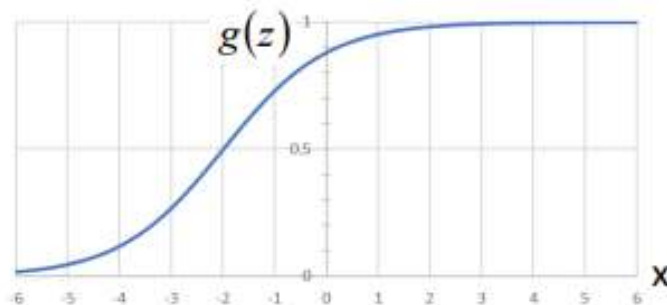
REGRESSIÓ LOGÍSTICA

La classificació es fa predint valors discrets, per exemple 0 o 1. De manera que es segueix una regressió lineal, però s'aplica un *threshold* que defineix a partir de quin punt les dades pertanyen a una classe o altra.

FUNCIÓ SIGMOIDE

$$g(z) = \frac{1}{1 + e^{-z}}$$

De manera que z és una funció. Si el pendent és gran, es comprimeixen les x .



HIPÒTESIS

$$h_{\theta}(x) = g(\theta^T X) = \frac{1}{1 + e^{-\theta^T X}}$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

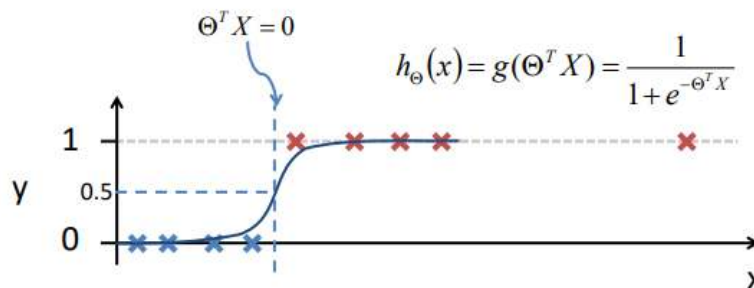
AC

2023-24

Llavors, és la probabilitat estimada que input x pertanyi a la classe $y=1$.

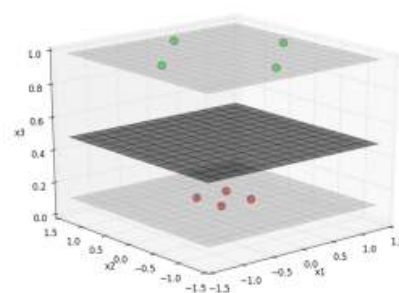
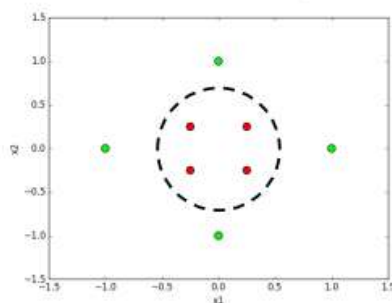
Decisió del límit

La decisió del límit està definida per $h_{\theta}(x) = 0.5$, que és equivalent a $\theta^T X = 0$.



Decisió límit no lineal

Si elevem al quadrat podem escriure un pla, podem separar el que era lineal. De manera que canviem de KERNEL.



WUOLAH

Funció de cost

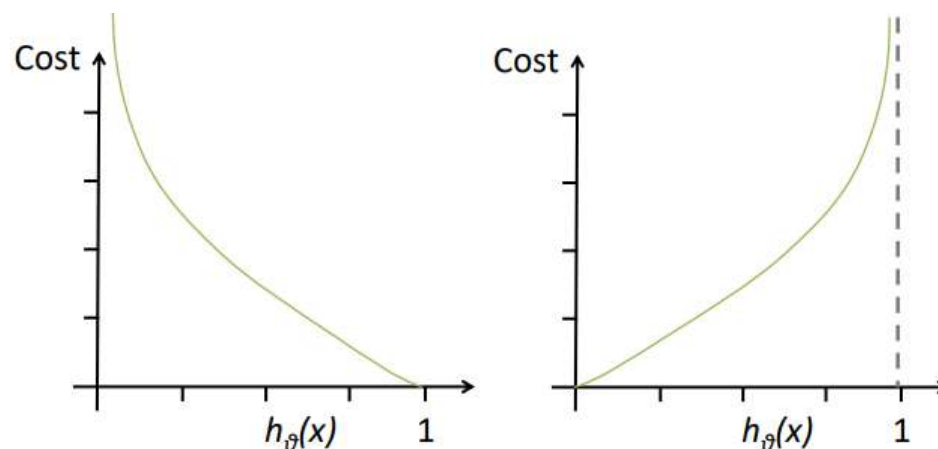
La funció de cost no és un convexa:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^i), y^i)$$

Si tenim que les funcions de cost per les classes 1 i 0 són:

- $-\log(h_{\theta}(x))$, if $y = 1$
- $-\log(1 - h_{\theta}(x))$, if $y = 0$

Si classifiquem bé les categories el cost serà 0 o gairebé. En canvi, si diem que un punt pertany a una classe que no és, estarem cometent un error **infinit**. A continuació tenim les gràfiques de cost per les classes 1 i 0, respectivament.



Si ajuntem les dues funcions de cost ens sortiria:

$$\text{Cost}(h_{\theta}(x), y) = -y \log(h_{\theta}(x)) - (1 - y) \log(1 - h_{\theta}(x))$$

Tenim en compte que y només pren valors 0 o 1.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

CLASSIFICACIÓ MULTI CLASSE

Tenim diversos mètodes per fer classificació quan tenim més de dues classes.

ONE VS ALL

El minimitzador triarà la classe més gran perquè serà amb la que s'equivocarà menys.

Amb c una classe, entrenem un classificador $h_{\theta}^c(x)$ per cada c per predir la probabilitat que $y_{\text{pred}} = c$. De manera que s'obté la probabilitat per cada classe i quina sigui més gran ens la quedem.

$$\arg_c \max h_{\theta}^c(x) \leftrightarrow \arg_c \max P(y = c|x, \theta)$$

ONE VS ONE

Si no es té molts recursos no es pot fer one vs one, tot i ser millor, ja que es fan totes les combinacions classe entre classe. Obtenim quants classificadors diuen que es pertany a una classe o es pot sumar les probabilitats.

Amb c una classe, entrenem un classificador $h_{\theta}^c(x)$ per cada parell de c . Utilitzant aquest mètode tindrem:

$$n \text{ classificadors} = nc * \frac{nc - 1}{2}$$

Exemple $nc = 10$

$$n \text{ classificadors} = 10 * \frac{10 - 1}{2} = 45 \text{ classificadors}$$

WUOLAH

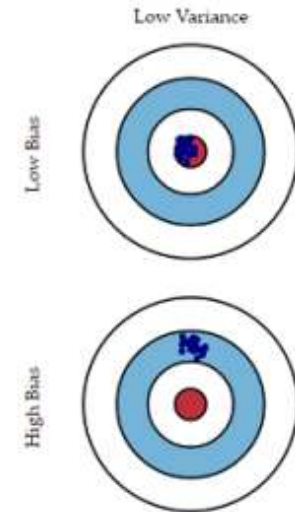
MODELS ERRONIS

UNDERFITTING

Pensem que el model és més simple del que és.

Els models “simples” tenen un alt error de **biaix** i baixa variància.

De manera que si repetim el *fit*, totes les solucions estaran d'esbiaixades cap una solució particular. Tanmateix, aquests models underfit són molt estables, de dimensions baixes i fàcils d'ajustar. També són fàcil de predir el seu comportament (perquè són esbiaixats).



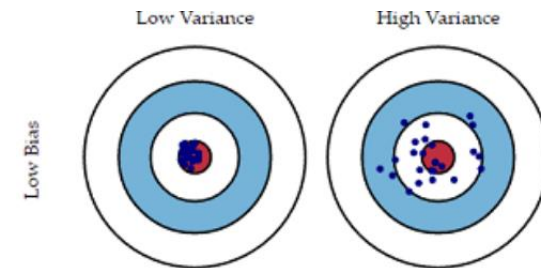
OVERFITTING

Ens equivoquem pensant que el model és més complex del que és. Aquests models tenen poc biaix i alta **variància**. De manera que si repetim el *fit* sobre diferents instàncies de problema, les solucions estaran per tot arreu.

Si tenim overfitting estem adaptant massa les dades d'entrenament.

La solució al model pot ser inestable, és a dir, no generalitza prou per les noves dades d'entrada que se li poden passar.

Es té una pèrdua mitjana molt baixa, s'ajusta amb poques dades i és útil quan estem segurs que la funció ideal precisa d'una explicació complexa.

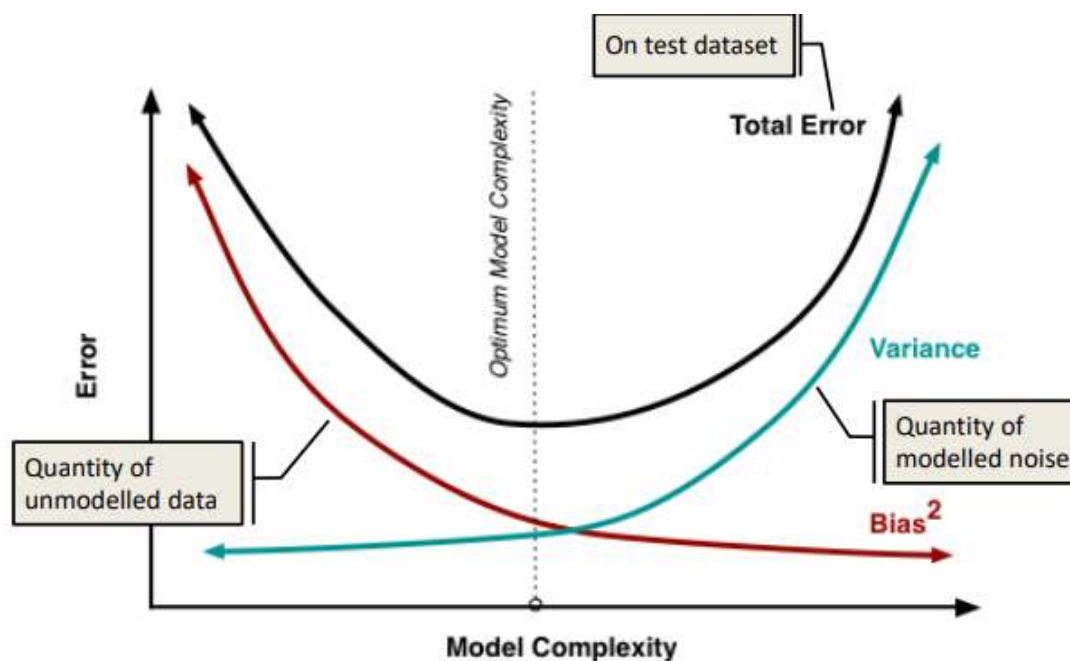


BIAIX

El biaix en un estimador és l'error resultant d'hipòtesis incorrectes fet al vostre model; l'aprenentatge pot perdre informació rellevant a causa del biaix.

VARIÀNCIA

La variància en un estimador és un error degut a la sensibilitat a fluctuacions en les dades; estimadors amb gran soroll d'ajust de variància i pot perdre la funció ideal subjacent.



Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

→ Plan Turbo: barato

→ Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

REGRESSIÓ LINEAL (underfitting)

Utilitzant la regressió lineal, afegim un paràmetre ϕ , que és la projecció de dades i perquè el model sigui més complex fem combinacions polinòmiques.

$$y \approx \phi(X^T)\theta$$

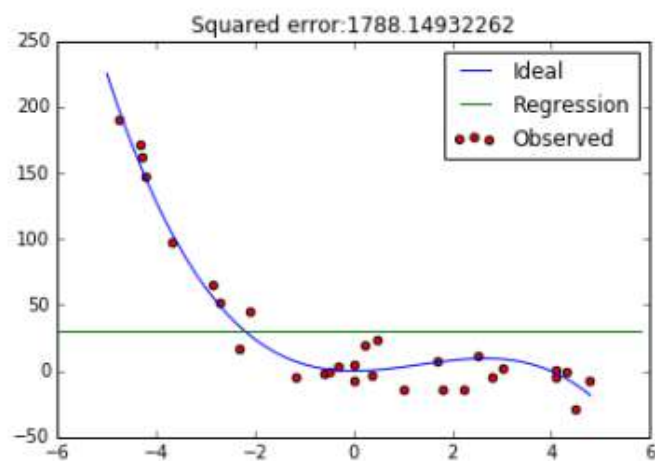
$$\hat{\theta} = \arg \min_{\theta} J(\theta; X, y, \phi)$$

$$J(\theta; X, y, \phi) = \sum_{i=1}^m L(\phi(x^i)^T \theta, y^i)$$

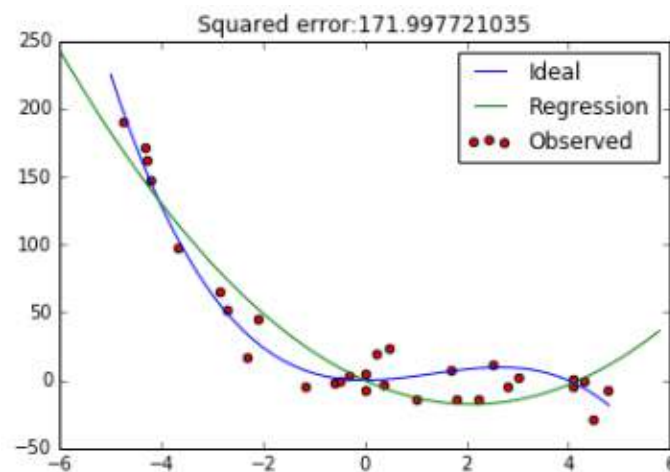
WUOLAH

Exemple

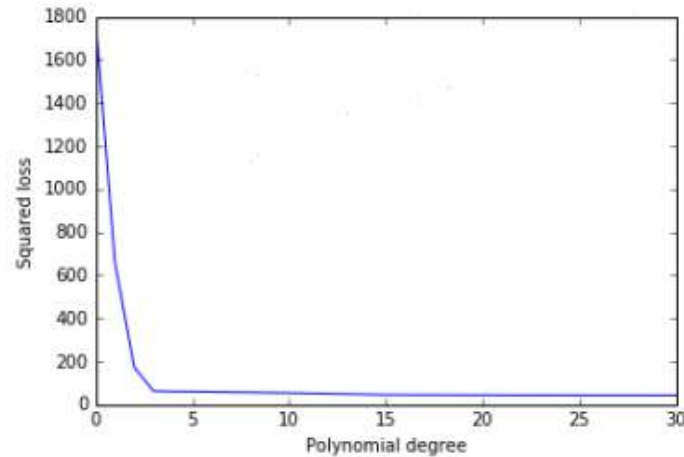
Amb $\varphi(x) = [1]$, veiem que l'error és molt alt, de manera que pugem el grau de la projecció fins que l'error sigui menor.



Augmentem el grau a $\varphi(x) = [1, x, x^2]$:



Per saber quan parar d'augmentar el grau de la nostra funció, si mirem la pèrdua el quadrat:



Mirant la gràfica de l'error parem d'augmentar el grau quan l'error s'estabilitza.

INSTÀNCIA DEL PROBLEMA

Considerem que les dades són generatives. De manera que per fer l'estimació d'una funció que té *underlying* farem servir mostres corruptes d'aquesta funció ideal f :

$$y^i = f(x^i) + N(0, \sigma)$$

On $N(0, \sigma)$ és soroll més o menys gaussià que li afegim a la funció, amb variància σ^2 :

$$N(0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

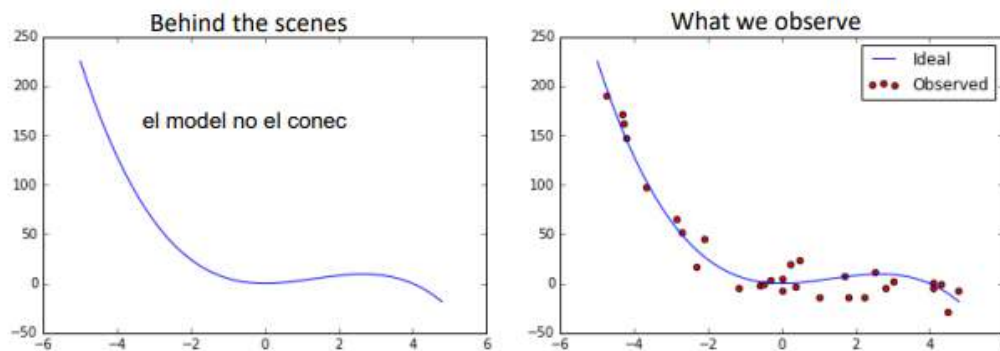
WUOLAH

AC

2023-24

Aquest soroll que afegim només el veiem a la funció y^i , no directament a f .

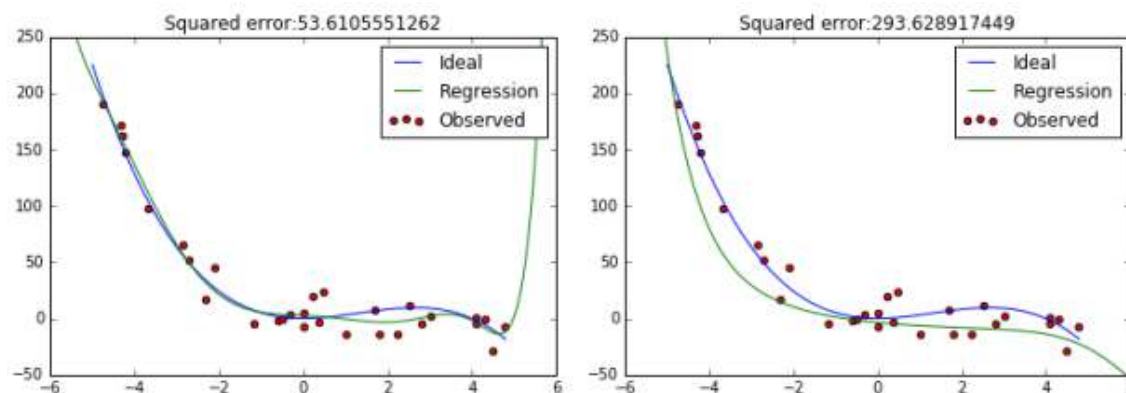
Exemple



WUOLAH

REGULARITZACIÓ

Comencem amb un **exemple**:



Tot i que el de l'esquerre té menor error quadràtic, el de la dreta és millor perquè s'equivoca més homogèniament. De manera que el model de l'esquerre té segurament *overfitting*, no generalitza.

Determinar la complexitat d'un model pot ser difícil, ja que no és simplement mirar el grau polinòmic o la dimensió. Per tant, en comptes de controlar la complexitat de la representació, sinó que ens fixarem en la complexitat d'ajustar el problema.

MAGNITUD DEL COEFICIENT

La magnitud del coeficient pot ser una bona mesura per la complexitat. És així com penalitzem els coeficients **regulant-los** (λ) a la funció de pèrdua.

$$\begin{aligned} J(\theta; X, y, \varphi) &= \sum_{i=1}^m L\left(\varphi(x^i)^T \theta, y^i\right) + \lambda(\theta^T \theta) = \\ &= \sum_{i=1}^m L\left(\varphi(x^i)^T \theta, y^i\right) + \lambda \sum_{k=1}^n (\theta_k^2) \end{aligned}$$

Com que es vol minimitzar el total, les dues parts han de ser-ho, però per donar més importància als pesos/coeficients escrivim lambda, anomenada:

- Regressió de Ridge
- l_2 ; regressió lineal regularitzada

Controla la compensació entre ajustar les dades (petita λ , alta variància) i minimitzar la complexitat del model (alta λ , baixa variància).

REGRESSIÓ DE RIDGE

Sense fer suposicions, pot ser difícil estimar el biaix i la variància dels nostres estimadors. En conseqüència, pot ser que no estigui clar quina λ triar. A més, els reguladors quadràtics donen solucions amb coeficients petits, però amb energia repartida uniformement (és a dir, són densos). Per tant,

- Si moltes dades, dividir en diversos plecs i fer servir la *cross* validació per estimar el model (i el biaix i la variància).
- Si necessitem una solució escassa, canviar el regulador.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

FUNCIÓ DE PÈRDUA LASSO – L_1

La funció de pèrdua Lasso (l_1), en comptes d'utilitzar el pes al quadrat, es fa servir l'absolut. Volem que els pesos siguin petits.

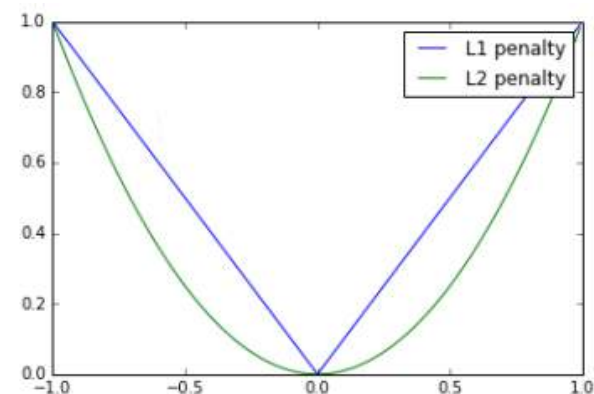
$$J(\theta; X, y, \varphi) = \sum_{i=1}^m L(\varphi(x^i)^T \theta, y^i) + \lambda \sum_{k=1}^n (|\theta_k|)$$

LASSO = Least Absolute Shrinkage and Selection Operator

Aquesta funció tria alguns coeficients no zero i tendeix a portar-los a 0.

LASSO VS RIDGE

Lasso	Ridge
l_1 fa possible que tots els models vagin de 1 a 0.	l_2 tendeix a fer valors petits.
Mai modifica els coeficients que tenen poc pes.	Penalitza els coeficients grans.
S'acosta bastant a l'ideal original.	El seu camí es troba enmig de les dades.
De tipus de selecció de model.	És una solució lineal densa.



WUOLAH

SVM

Abans de començar recordem:

Models generadors

S'intenta modelar la classe. Estimen què fan les variables que fluctuen.

S'aprèn $P(X|y)$

Models discriminadors

Utilitza les dades per trobar la millor manera per separar les classes.

S'aprèn: $P(Y|X)$

És més difícil tenir un model generador que un descriptiu, ja que el generador explica el comportament de les dades.

SVM o Suport de Vectors de Màquines té les components:

- **Classificadors lineals:** el límit de decisió és un hiperplà. Utilitzen una recta per separar les dades en classes.
- **Màxim marge:** la solució maximitza el marge (on no hi ha punts/dades dins) entre les classes. Es vol un marge el més gran possible.
- **Vectors de suport:** uns quants punts de les dades són necessaris per definir el marge.

El marge **canvia** quan agafem altres decisions com a límit (el marge és més estret), o si els vectors de suport es mouen. En canvi, si la resta dels punts es mouen, el marge es manté igual.

A més, en comparació amb la Regressió Logística, SVM output no es pot tractar com una probabilitat.

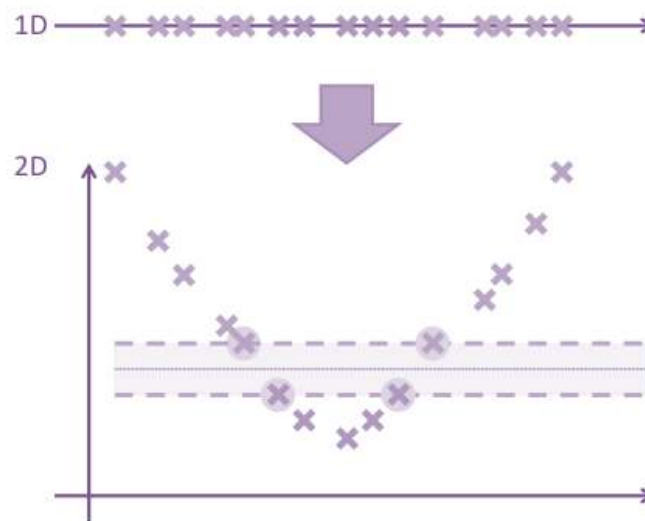
OUTLIERS

Si els outliers són els vectors de suport pot afectar la solució greument. En aquest cas es pot relaxar la condició de màxim del marge (*soft margin*). És a dir que es deixa que hi hagi algun punt dins de la frontera.

Això implica afegir tolerància als errors, controlada per les variables *slack*. Aquestes variables introdueixen una regularització i per tant, relaxen la condició de marge màxim, és a dir, l'error.

KERNEL TRICK

Si les classes no són linealment separables, SVM no serà capaç de treballar. Llavors, podem canviar la dimensió del problema, canviar el nucli.



Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

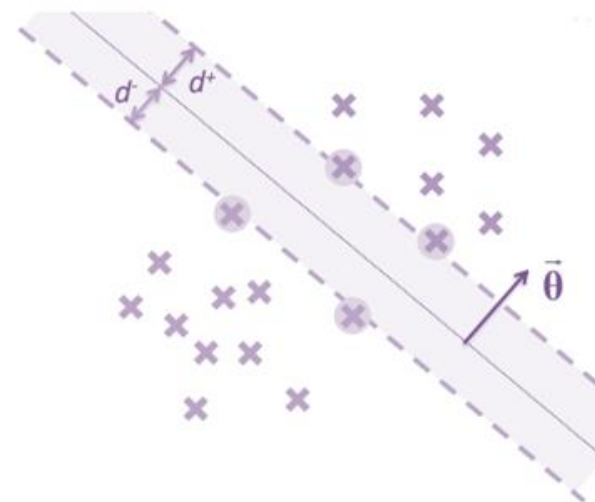
AC

2023-24

OBJECTIU

El nostre objectiu és trobar l'hiperplà que maximitzi la distància entre els vectors més proper (de les dues classes). Equivalent a maximitzar el marge.

- $\vec{\theta}$ representa el vector que defineix l'hiperplà.
- d^+ és la distància de l'hiperplà als vectors de suport de la classe, en aquest cas (+).
- d^- és la distància de l'hiperplà als vectors de suport de la classe, en aquest cas (-).
- $D = |\mathbf{d}| = d^+ = d^-$



Per maximitzar:

$$D = \frac{\vec{\theta} * \vec{x} + \theta_0}{|\vec{\theta}|} = \frac{1}{|\vec{\theta}|}$$

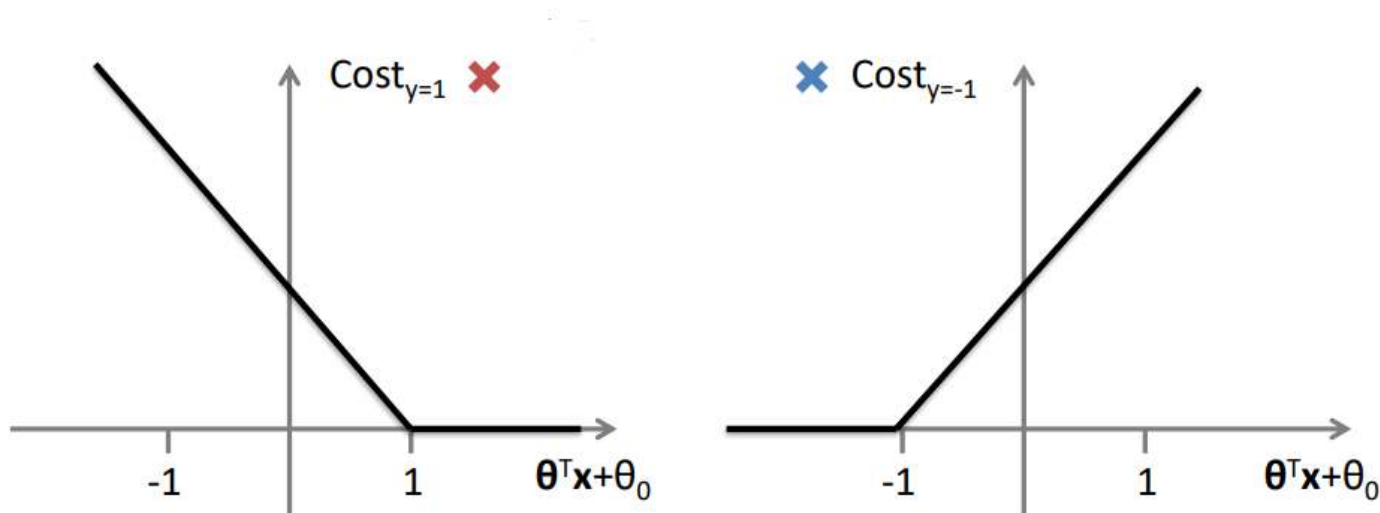
De la funció següent, si tots els punts estan ben classificats, l'error = 0, de manera que s'intenta minimitzar el segon terme, que farà possible maximitzar el marge.

$$\min_{\theta} C \sum_{i=1}^m [y^{(i)} \text{cost}_1(\theta^T \mathbf{x}) + (1 - y^{(i)}) \text{cost}_0(\theta^T \mathbf{x})] + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

WUOLAH

FUNCIONS DE COST

Per calcular el cost no penalitzem el que està classificat correctament, però sí els punts que es troben en el costat equivocat. A les gràfiques següents, en el cas que tinguem dues categories (1 i -1), podem veure que el cost és alt quan classifiquem malament i que com més lluny de la frontera més cost.



REPRESENTACIÓ DUAL

En representació dual busquem els punts de suport, no theta ni hiperplà.

$$\left. \begin{array}{l} \min_{\theta, \theta_0} \Phi(\theta) = \frac{1}{2} \theta^T \theta \\ \text{Subject to:} \\ y_i(\theta^T x_i + \theta_0) \geq 1 \end{array} \right\} \begin{array}{l} \text{P} \\ \max_{\alpha} (\min_{\theta, \theta_0} (L(\theta, \theta_0, \alpha))) \end{array}$$

SVM PRIMAL

$$\begin{array}{l} \max_{\alpha} \Phi(\alpha) = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i \\ \text{Subject to: } \sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0 \end{array}$$

SVM DUAL

El problema dual permet obtenir una solució explícita a partir de productes escalars.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

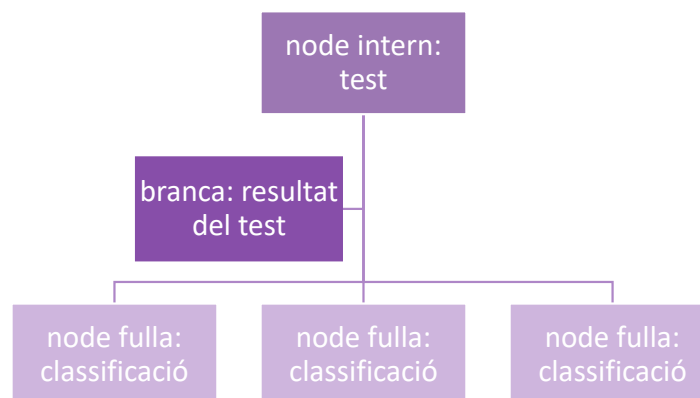
AC

2023-24

ARBRES DE DECISIÓ

Hi ha casos que no podem aplicar el descens del gradient. Per això introduïm el classificador: **arbre de decisió**.

Aquest classificador permet processar una entitat que ha estat descrita per un conjunt de propietats i prendre una decisió. Té els components:



Si tenim dos classificadors, un amb moltes fulles i l'altra amb no tantes: es considera que el bon classificador és el que té menys.

Binari

Pel cas de classificació binària, poden representar qualsevol funció. I la taula de veritat d'una funció booleana de n característiques té 2^n files (x), i per tant hi ha 2^x funcions booleanes diferents.

WUOLAH

CONSTRUCCIÓ DELS ARBRES

Mètode recursiu

1. Si resten exemples positius i negatius, escollir el millor atribut per dividir-los.
2. Si tots els exemples són positius o negatius ja hem acabat i podem respondre.
3. Si no hi ha exemples vol dir que no hem observat cap exemple d'aquest tipus: retornem un valor per defecte calculat a partir de la classificació majoritària del node pare.
4. Si no hi ha atributs i queden exemples positius i negatius tenim un problema: els exemples tenen la mateixa descripció però diferent classificació. Hi ha soroll a les dades!

Criteris de parada

- Quan tots els exemples que quedin pertanyin a la mateixa classe (s'afegeix una fulla a l'arbre amb l'etiqueta de la classe).
- Quan no quedin atributs pels que ramificar (s'afegeix una fulla etiquetada amb la classe més freqüent en el node).
- Quan no quedin dades per classificar.

HEURÍSTIQUES

Per ramificar un arbre o per seleccionar els atributs podem aplicar diverses heurístiques, la millor serà que ens proporcionin nodes més homogenis. De manera que si no generalitza, tenim un model dolent.

Exemples: Guany d'informació (ID3, C4.5), índex de Gini (CART, SLIQ, SPRINT)

Guany d'informació

Algorisme ID3

Aquest algorisme utilitza el criteri de divisió de **relació** del guany d'informació basat en l'ENTROPIA, que mesura la quantitat de desordre.

$$Entropia(S) = - \sum_{x=1}^m p(x) \log_2(p(x))$$

On $p(x)$ és la probabilitat que un exemple S pertanyi a classe i .

Per calcular l'entropia després de fer una subdivisió:

$$Entropia(S, A) = \sum_{v \in A} \frac{|S_v|}{|S|} Entropia(S_v)$$

On S_v és el nombre de respostes/total, $Entropia(S_v)$ és la suma de les entropies de les possibles respostes (dins del subconjunt).

Així, divideixo en tants subconjunts com tipus que té l'atribut escollit, per exemple, "tens gana?" si o no, llavors partim en 2. Si necessitem pocs bits, llavors tenim pocs objectes a classificar. Millor tenir 1 bit que 8 bits.

$$Gain(S, A) = Entropia(S) - Entropia(S, A)$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo
espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

Com més gran la diferència del guany, és a dir, que com més gran el guany obtingut fruit de la partició, millor i llavors triarem aquest classificador.

Algorisme C4.5

El criteri de divisió és la **proporció** del guany (*gain ratio*, C4.5).

En comptes de ramificar l'arbre utilitzant els atributs que tenen més valors diferents, com fa ID3, fet que normalitza el guany d'informació amb l'entropia. Per això, aquest algorisme busca l'entropia sobre el propi valor (no dins de la classe).

$$SplitInfo(S, A) = - \sum_{v \in A} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$$

Ens interessa que la quantitat repartida entre branques sigui bastant igual i llavors, tingui un guany homogeni entre branques.

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

WUOLAH

Índex de Gini

Algorismes CART, SLIQ, SPRINT

Es fa servir una mesura estadística d'impuresa. I és proporcional amb l'entropia, quan augmenta Gini, també ho fa l'entropia.

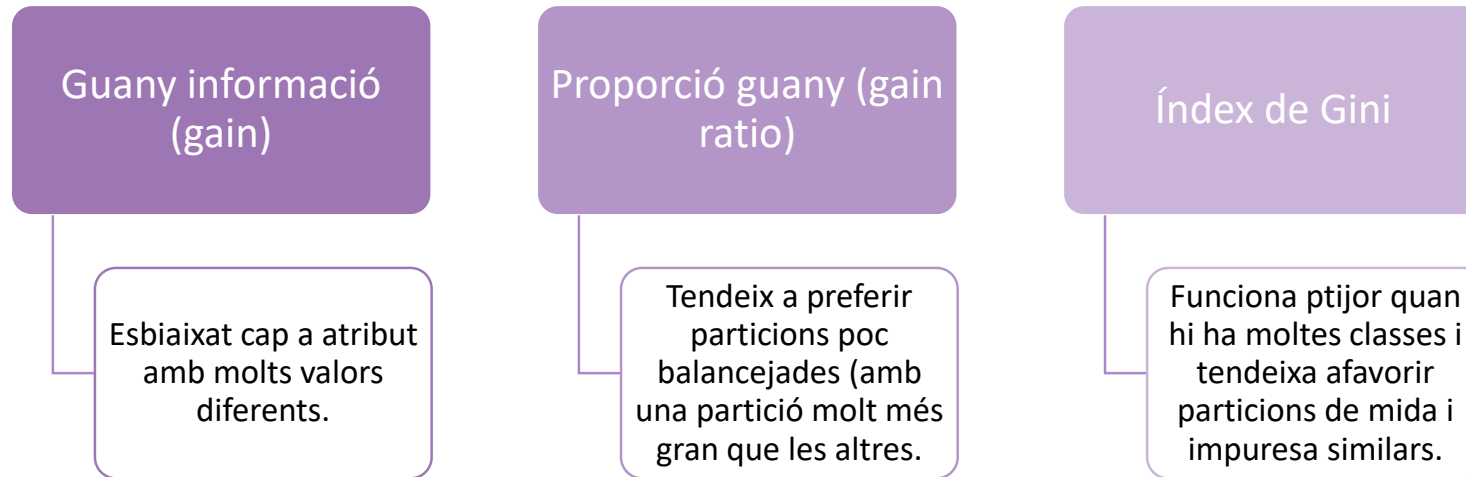
$$Gini(y, S) = 1 - \sum_{v \in \text{dom}(y)} p(v|S)^2 = 1 - \sum_{v \in \text{dom}(y)} \left(\frac{|S_v|}{|S|} \right)^2$$

I la Gini després de dividir:

$$GiniGain(S, A) = Gini(y, S) - \sum_{v \in A} \frac{|S_v|}{|S|} Gini(y, S_v)$$

Per a construir l'arbre, triem l'atribut que proporciona major reducció d'impuresa.

Comparativa criteris de divisió



PROBLEMES

Per corregir *overfitting* s'ha d'eliminar branques (podar). De manera que, un cop s'ha après: podar.

Podem tenir un rendiment baix o bé per la presència de soroll a les dades (mal etiquetades) o bé, pocs exemples en relació a la complexitat de la funció. Per solucionar-ho, ens aturem abans de classificar perfectament les dades o post-processar l'arbre: pruning.

Si tenim variables contínues les podem discretitzar.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

PODA

Tots els nodes de decisió són candidats a podar-se. Un node és eliminat només si l'arbre resultant no és pitjor que l'anterior sobre el conjunt de validació. Llavors, eliminar el subarbre que penja del node, convertir-lo en fulla, i assignant el valor majoritari dels exemples que hi pengen com a valor de classificació.

Reduced-Error Pruning

Poda per reducció d'error

Classificar els exemples en el conjunt de validació – alguns podran ser erronis. Per a cada node:

- Sumar els errors en tot el subarbre.
- Calcular l'error amb els mateixos exemples si es converteix el subarbre en una fulla amb l'etiqueta majoritària.
- Podar el node amb la major reducció d'error.
- Repetir fins que l'error no es redueix.

Pruning Pesimistic

Evita usar el conjunt d'avaluació (podrem entrenar en més exemples). També estima de forma conservadora l'error real a cada node, basant-se en els exemples d'entrenament. "Correcció de continuïtat" a la taxa d'error a cada node: afegim $1/2N$ als errors observats, on N es el número de fulles en el sub-arbre. I podem el node a menys que l'error estimat del sub-arbre és mes petit de 1 error standard per sota del error estimat per al podat: $r'_{\text{subtree}} < r'_{\text{pruned}} + SE$

WUOLAH

Cost-complexity Pruning

És una mesura de l'error promig reduït per fulla. Calcula la mesura α per a cada node si es substitueix per una fulla. I compara els errors en les fulles, i podar el de mínima α .

S'itera fins que no hi ha cap node amb $\alpha < \text{llindar}$.

CONJUNTS DE MÈTODES

Utilitzar diferents dades d'aprenentatges o algorismes, de manera que creem diferents respostes amb el conjunt de dades, a partir de les quals podem combinar les decisions. Estem sumant coneixements.

Tot i que es pot combinar diferents dades o mètodes, normalment es fa servir el mateix arbre de decisió però amb diferent dades, com el que es coneix *Ensembles* homogenis.

Avantatges

Quan combinem múltiples decisions independents i diverses cada una de les quals és millor que l'atzar, els errors deguts a atzar es cancel·len els uns als altres, i les decisions correctes es reforcen.

ENSEMBLES HOMOGENIS

Com hem definit abans, un conjunt homogeni és quan utilitzem un algorisme d'aprenentatge arbitràriament i només manipulem les dades per fer-lo aprendre múltiples models.

INCÍS

Si parlem de dades heterogènies, ens referim a fer servir diferents classificadors amb diferents dades.

Mètode per canviar dades

BAGGING

Re-mostregem les dades d'aprenentatge. De manera que primer utilitzem una part i llavors un altra.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

→ Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

BOOSTING

Re-pesem les dades d'aprenentatge.

DECORATION

Afegim dades d'aprenentatge addicionals artificials.

Minimum distance classifier

És un mètode que classifica vectors característics basat en seleccionar la distància més curta d'un vector d'entrada respecte a tots els centres de classe i assignant el vector al centre de classe que revela la distància més curta. Per tant, es mira el punt entremig entre els dos punts.

Decision Boundaries

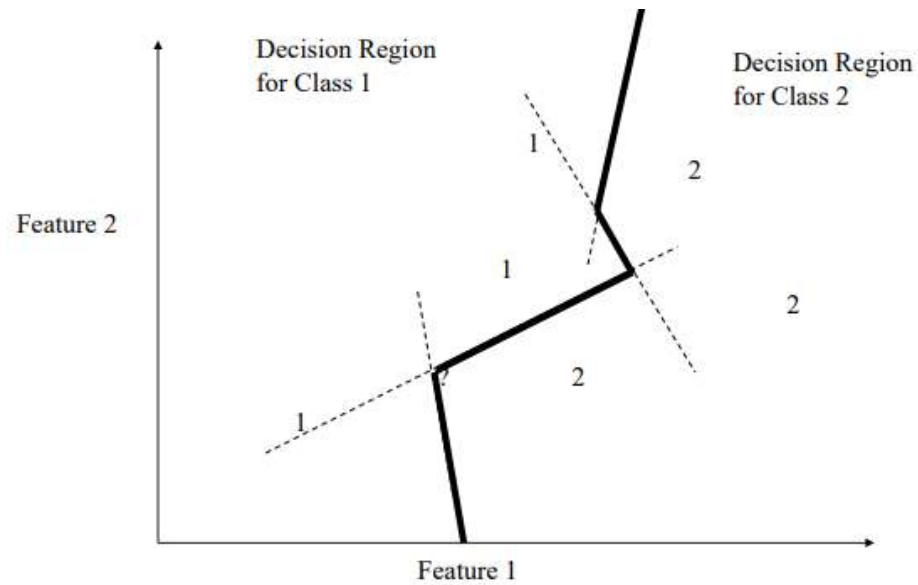
S'agafen dos punts equidistants entre els punts de la classe 1 i 2, i es traça línia (límit). Localment el límit és lineal.

Overall Boundary = Piecewise Linear

S'ajusta massa al training, fet que comporta *overfitting*.

WUOLAH

Exemple overall boundary:



BAGGING AND BOOSTING

Predictors/classificadors d'agregació construïts a partir de reponderades versions del conjunt d'aprenentatge.

Bagging

Bagging = **Bootstrap Aggregating**

Tenint n dades, es tria aleatòriament dades del conjunt original n vegades (amb la possibilitat de repetir-se mostres).

El *reweightning* dels conjunts d'aprenentatge es fa dibuixant a l'atzar amb substitució dels conjunts d'aprenentatge.

Els predictors s'agreguen per votació plural, per exemple un element e s'ha classificat dues vegades 0 i una vegada 1, per tant, $e = 0$.

Es pot fer servir **suma de les sortides**

$$c_{bag(x)} = \text{sign} \left(\frac{1}{B} \sum_{b=1}^B c^b(x) \right)$$

o **suma de les probabilitats**.

$$p_{bag(x)} = \frac{1}{B} \sum_{b=1}^B p^b(x)$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

→ Plan Turbo: barato

→ Planes pro: más coins

perdo
espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

Boosting

Boosting es refereix a un mètode general que produeix una regla de predicció molt precisa combinant *rough* i moderadament imprecisa regles generals.

Tenim classificadors petits i en procés d'aprenentatge es mira si s'ha equivocat. Llavors el classificador següent mira l'avaluació d'entrenament del classificador anterior i es donarà més importància en les observacions que s'ha equivocat: les dades són remostrejades (resampeld) de manera adaptativa.

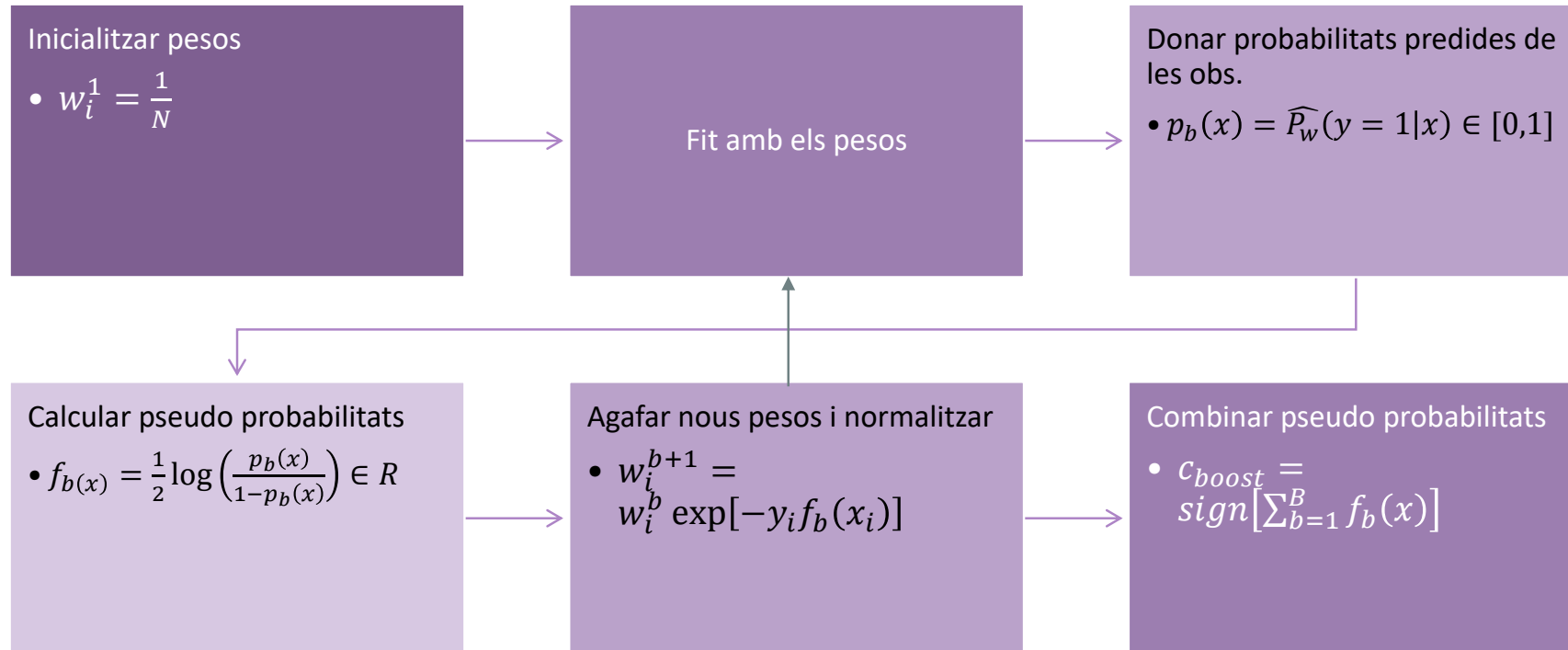
$$Error = w(y_{test} - y_{pred})^2$$

On w és el pes que se li afegeix quan el classificador anterior s'ha equivocat.

AdoBoost

Se li pot aplicar e Descens del Gradient (Gradient Boosting).

WUOLAH



RANDOM FOREST

Random Forest fa créixer molts arbres. De manera que el bosc és un *ensemble* d'arbres de decisió no podats. Cada classificador base, classifica un nou vector d'atributs de les dades originals. I finalment, el 'bosc' tria la classificació resultant que ha tingut més vots (de tots els arbres del bosc).

Utilitza *bagging* i vector d'entrada aleatori. El primer, cada arbre creix usant una mostra 'bootstrap' de les dades d'aprenentatge. I el segon, a cada node, la millor divisió s'escull entre una mostra aleatòria de m atributs enlloc d'entre tots els atributs.

Per triar el nombre de mostres a utilitzar en cada selecció aleatòria, ens guiem per:

m	si
$\frac{M}{3}$	Regressió
$ \sqrt{M} $	Classificació
" <i>tunning paràmtre</i> "	Ajustar el paràmetre

ALGORISME

Sigui N el número de casos d'aprenentatge, i M el número de variables de les mostres.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo
espacio

AC

2023-24

Fixem un número m de variables d'entrada per usar en el test de la decisió en un node de l'arbre; m hauria de ser bastant inferior a M .

Escollim un conjunt d'aprenentatge per a un arbre, escollint n vegades amb reposició d'entre els N casos d'aprenentatge disponibles (i.e. Prenem una mostra bootstrap). Usem la resta dels casos per a estimar l'error del arbre, predint les seves classes.

Cada arbre creix fins al límit i NO es poda.

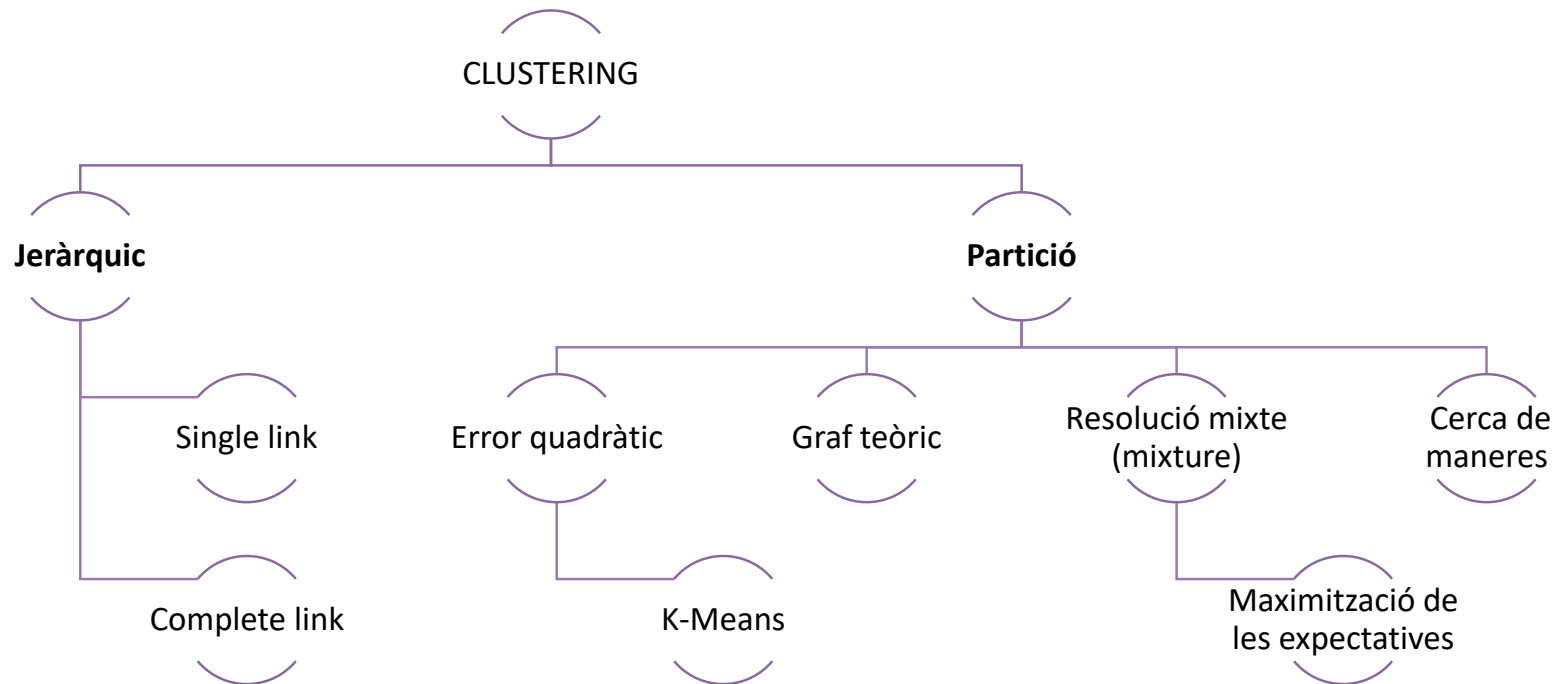
Per a cada node de l'arbre, aleatoriament triem m variables en base a les que prendrem la decisió en aquest node. Calculem la millor divisió d'aquestes m variables del conjunt d'aprenentatge.

WUOLAH

CLUSTERING

És un model no supervisat, de manera que no tenim un *groundtruth*. I busca agrupar objectes semblants entre ells i diferent amb els altres, pertanyent a un altre grup.

TIPUS DE CLUSTERS



Clustering partitional

Es fa un divisió de les dades en subconjunts (clústers) no solapats. Per això, un objecte pertany a un sol grup.

K-Means

Per portar a terme un algorisme de K-Means:

1. Inicialitzar K "significa" μ_k , un per a cada classe.
2. Assignar cada punt a la mitjana més propera μ_k .
3. Actualitzar els mitjans μ_k dels nous clústers.
4. Si els mitjans no canvien acabar, sinó tornar a **pas 2**.



Té problemes quan els clústers tenen diferent mida, densitat i/o no tenen forma circular. A més, els *outliers* també causen errors. I els punts només poden pertànyer a una classe.

Tria del nombre de clústers

A l'hora de triar els punt inicials, la probabilitat que K sigui un centroides és relativament petita quan K és gran, si els clústers són de la mateixa mida, n, llavors:

$$P = \frac{\# \text{maneres de seleccionar centroides de cada clúster}}{\# \text{maneres seleccionar } K \text{ centroides}} = \frac{K!}{K^K}$$

Exemple

Si K = 10:

$$P(K = 10) = \frac{10!}{10^{10}} = 0.00036$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

Els primers clústers afegiran molta informació (explicar molta variància / reduir molt el valor de la funció de cost), però en algun moment el guany marginal baixarà, donant un angle en el gràfic. El nombre de clústers es tria en aquest punt, per tant "elbow criterion". Aquest "colze" no sempre es pot identificar sense ambigüitats.

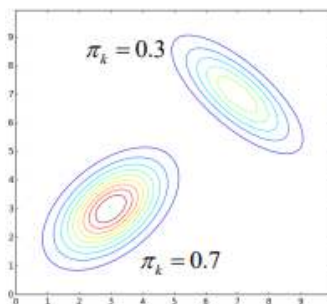
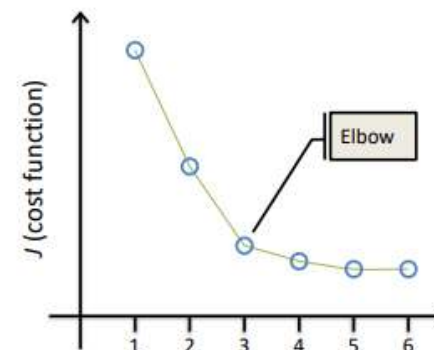
$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \left\| \mathbf{x}^{(n)} - \boldsymbol{\mu}_k \right\|^2$$

On $r_{nk} \in \{0,1\}$, μ_k és el vector assignat a cada punt de les dades.

Gaussian mixture models

És un model probabilístic i és *soft*, ja que un punt pot pertànyer a diverses classes amb probabilitat k .

Si el punt està a prop del centroide es considerarà de la classe en qüestió.



WUOLAH

Densitat

$$p(\mathbf{x}) = N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Mixtures

$$N(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + N(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

$$p(\mathbf{x}) = \pi_1 N(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \pi_2 N(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

Variable latent

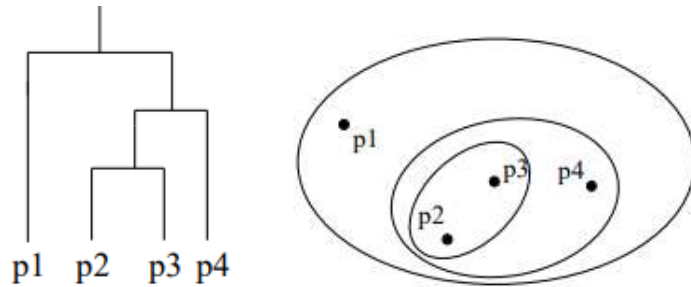
La variable latent z és una variable aleatòria binària K -dimensional amb una representació 1 de K ,

Clustering jeràrquic

És un algorisme que pot ser:

- Aglomeratiu (*down-up*): es comença tots els nodes formant un clúster amb ells mateixos i es va ajuntant.
- Divisiu (*up-down*): es comença per un clúster compost per tots els nodes i a cada pas es va dividint, generant clústers nous.

Per visualitzar el procés ho podem fer a través del dendrograma.



Amb comparació als models anteriors, no s'ha de assumir cap número de clústers.

Single link (min)

La similitud de dos clústers es basa en els dos més semblants punts (més propers) dels diferents clústers.

Pot manejar formes no el·líptiques, però és sensible al soroll i als valors *outliers*.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

→ Plan Turbo: barato

→ Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

Complete linkage (max)

La similitud de dos clústers es basa en els dos menys semblants (la majoria distants) en els diferents grups.

Aquest mètode és menys susceptible al soroll i als valors *outliers*. Tanmateix, tendeix a trencar grans grups i és esbiaixat cap als clusters globulars.

Group average

La proximitat de dos clústers és la mitjana de la proximitat per parelles entre punts dels dos grups.

Es troba entre single link i complete link, ja que:

- Menys susceptible al soroll i als valors atípics.
- Esbiaixat cap als cúmuls globulars.

Ward's Method

La semblança de dos grups es basa en l'augment del quadrat error quan es fusionen dos clústers

- Similar a *Group average* si la distància entre punts és distància al quadrat.
- Anàleg jeràrquic de K-means.

Es pot utilitzar per inicialitzar K-means.

WUOLAH

VALIDACIÓ DEL MÈTODE CLUSTER

1. Determinar la tendència d'agrupació d'un conjunt de dades, és a dir, distingir si realment existeix una estructura no aleatòria a les dades.
2. Comparar els resultats d'una anàlisi de clústers amb els resultats coneguts externament, per exemple, amb etiquetes de classe donades externament.
3. Avaluar fins a quin punt els resultats d'una anàlisi de clústers s'ajusten a les dades sense fer referència a informació externa.
 - Utilitza només les dades.
4. Comparar els resultats de dos conjunts diferents d'anàlisis de clústers per determinar quin és el millor.
5. Determinació del nombre 'correcte' de clústers.

MESURES INTERNES

Cohesió del clúster: mesura la relació estreta que hi ha entre els objectes a clúster. Exemple: SSE

$$WSS = \sum_l \sum_{x \in C_l} (x - m_l)^2$$

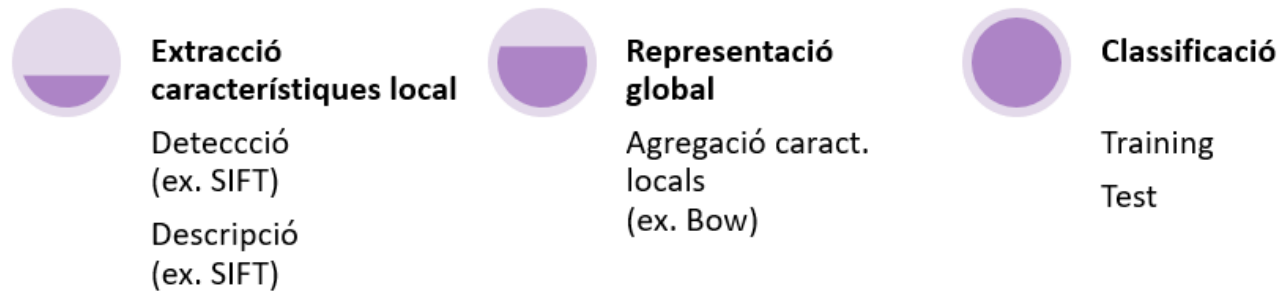
Separació de clústers: mesura com de diferent o ben separat a clúster és d'altres clústers. Exemple: error quadrat

$$BSS = \sum_l |C_l| (m - m_l)^2$$

BAG OF WORDS

En aquest model no es té en compte la distribució, amb els histogrames veiem la quantitat de vegades qu apareix cert valor. El resultat de fer un Bag of Words és: reducció de la sesibilitat al soroll i increment de la capacitat de discriminació.

En el cas de les imatges es fan servir els descriptors locals, si aquests s'assemblen, les imatges pertanyen a la mateixa classe. SIFT és un mètode extractor de descriptors.



Els detectors de característiques estan dissenyats per trobar característiques locals invariants que permeten trobar una correspondència entre imatges similars. Aquestes característiques locals no són necessàriament rellevants per a la classificació. I no proporcionen una cobertura total de tota la imatge, una propietat que pot ser important per a la classificació.

Importante

pierdo
espacio



Necesito
concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

Puedo eliminar la publi de este
documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

AC

2023-24

DENSE SAMPLING

Donar/agafar una característica en interval de píxels o temps. De tal manera que cobrim tota la imatge i podem icloure informació sobre el context dels objectes.

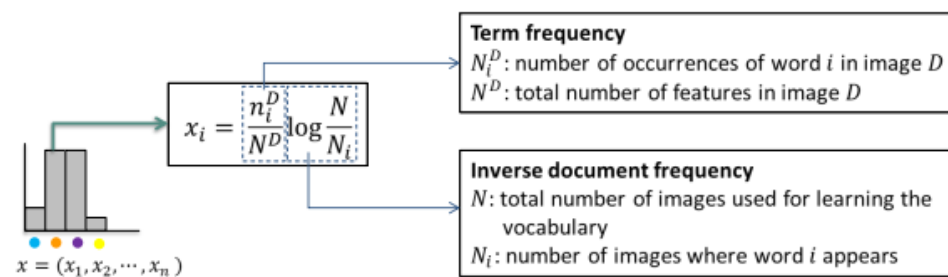
Com més punts claus, més computació i memòria necessària. A més, tenint masses punts pot fer que es tinguin punts que no proporcionin informació.

REPRESENTACIÓ DE L'ESTÍMUL

Term frequency – Inverse document frequency (tf-idf)

A cada “paraula” se li assigna un pes/importància, a través de:

- Freqüència del terme - tf: nombre d'ocurrències de la paraula visual a la imatge (normalitzada pel nombre total de característiques de la imatge).
- Freqüència inversa del document - idf: assignar un pes més gran al paraules visuals més significatives (paraules menys freqüents en el conjunt general d'imatges).



WUOLAH

CLASSIFICACIÓ

Es pot fer servir SVM, funció de Kernel...

SVM multiclasse

One vs one

De totes les classes que es té es fa un classificador 1-1. Després aplicar-ho a tots els classificadors i votar a la classe guanyadora de cada classificador.

Es fan servir $n * \frac{n-1}{2}$ classificadors diferents.

Resultat de la classificació: classe amb més vots.

One vs all

Entrena un classificador 1-n per a cada classe amb mostres de totes les altres classes com a mostres negatives.

Es fan servir n classificadors diferents.

Per classificar, aplicar tots els classificadors i obtenir la puntuació de classificació normalitzada.

Resultat de la classificació: classe amb una puntuació de classificació normalitzada més gran.

Kernel

En el cas dels **kernels lineals**, són més eficients tant en el conjunt d'entrenament com el de test:

$$K(\mathbf{x}, \mathbf{z}) = \langle \mathbf{x}, \mathbf{z} \rangle \longrightarrow \left. \begin{array}{l} f(\mathbf{x}) = \text{sgn} \left(\sum_i y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \\ \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \end{array} \right\} f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$

Els **kernels no lineals** solen ser donar millors classificacions:

Gaussian (RBF): $K(\mathbf{x}, \mathbf{z}) = e^{-\gamma \ \mathbf{x} - \mathbf{z}\ _2^2}$	} $f(\mathbf{x}) = \text{sgn} \left(\sum_i y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \right)$
Intersecció: $K(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \min(\mathbf{x}(i), \mathbf{z}(i))$	
Chi-squared(X^2): $K(\mathbf{x}, \mathbf{z}) = \frac{2\mathbf{xz}}{\mathbf{x} + \mathbf{z}}$	
Hellinger (Bhattacharya): $K(\mathbf{x}, \mathbf{z}) = \sqrt{\mathbf{xz}}$	Explicit mapping function $\varphi(\mathbf{x}) = (\sqrt{x_1}, \dots, \sqrt{x_n})$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

TIPUS D'ASSGINACIONS

Hard assignment	Soft assignment
Cada característica local s'assigna a paraula visual més propera.	S'assigna cada característica local a diferents paraules visuals.
L'aportació de cada local característica a cada <i>bin</i> de l'histograma és 1/0.	L'aportació de cada local característica a cada <i>bin</i> de l'histograma és una funció de la distància de la característica al centre del clúster.

VARIABLES LATENTS

Espai latent són les variables que no estan indicades però que es tenen en compte.

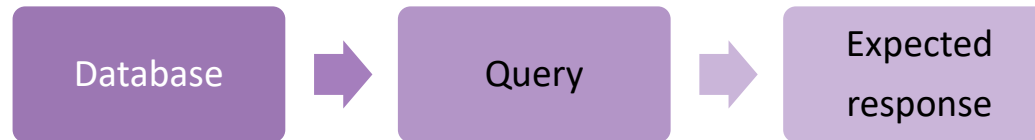
WUOLAH

SISTEMES RECOMANADORS

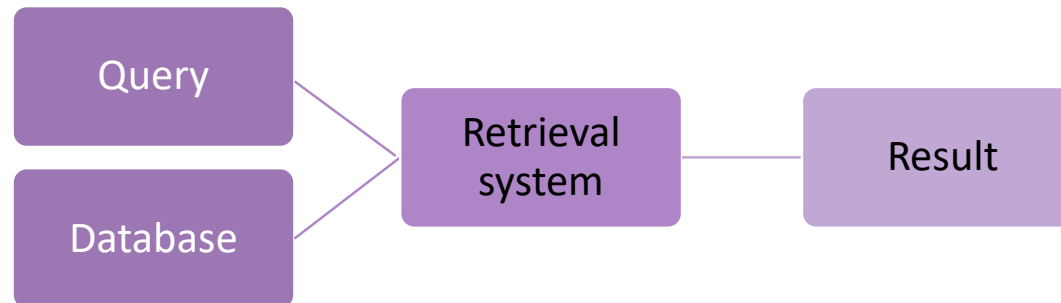
MESURA DE LA RELLEVÀNCIA

La metodologia estàndard en l'avaluació del rendiment de la recuperació d'informació consisteix en tres elements:

- Una col·lecció d'articles de referència
- Un conjunt de consultes de referència
- Una avaluació de la rellevància de cada parella consulta-element (cada element és rellevant o no a la consulta).



OBTENCIÓ D'INFORMACIÓ



A partir de la informació podem construir una matriu de confusió, on:

- $|rel|$ vol dir rellevant, coincident amb la *query*
- $|\overline{rel}|$, no rellevant
- $|ret|$, obtingut
- $|\overline{ret}|$, no obtingut, que estan al *database*

	Relevant	Non-relevant	TOTAL
Retrieved	$ ret \cap rel $	$ ret \cap \overline{rel} $	$ ret $
Non retrieved	$ \overline{ret} \cap rel $	$ \overline{ret} \cap \overline{rel} $	$ \overline{ret} $
TOTAL	$ rel $	$ \overline{rel} $	

De la matriu anterior podem calcular la precisió i *recall*.

Precisió

És la fracció dels documents obtinguts (després de fer query) que són rellevants. Mesura la qualitat del sistema de retorn (*Retrieval System*), és a dir, la seva capacitat d'incloure ítems rellevants al resultat.

$$Precision = \frac{|ret \cap rel|}{|ret|} = \frac{\# \text{ relevant items retrieved } (= TP)}{\# \text{ retrieved items } (= TP + FP)} = P(\text{relevant}|\text{retrieved})$$

Recall

És la fracció dels documents rellevants que s'han retornat. Mesura l'eficiència del sistema en retornar tots els elements rellevants.

$$Recall = \frac{|ret \cap rel|}{|rel|} = \frac{\# \text{ relevant items retrieved } (= TP)}{\# \text{ relevant items } (= TP + FN)} = P(\text{relevant}|\text{relevant})$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

F-measure

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}}$$

NOTA Amb els Recomenadors no es fa servir l'*accuracy*, ja que és la fracció dels rellevants+no-rellevants que són correctes.

De manera que obtindríem una bona *accuracy* sempre, quan només ens interessa el que es retorna sigui rellevant.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

REALITZACIÓ RANKING

Per ordenar com de segur estem fem servir predict_proba.

	P	R	
1	1/1	1/9	@1
1	2/2	2/9	@2
0	2/3	2/9	@3
1	3/4	3/9	@4
0	3/5	3/9	@5
0	3/6	3/9	@6
1	4/7	4/9	@7
1	5/8	5/9	@8
...

Total Relevant in the dataset: 9

WUOLAH

PRECISIÓ MITJANA

Average Precision premia el retorn més aviat d'elements rellevants. Ranking és important.

Recuperar tots els elements rellevants de la col·lecció i classificar-los perfectament donarà una precisió mitjana d'1.

Per tant, la precisió mitjana de la precisió mitjana en tots els valors de *recall* entre 0 i 1. Això és igual a l'àrea sota la corba P-R. La corba perfecte seria tot el gràfic, amb àrea = 1.

$$AP = \int_0^1 p(r) dr = \sum_{k=1}^N P(k) \Delta r(k) = \frac{\sum_{k=1}^N P(k) rel(k)}{|rel|}$$

De l'exemple anterior:

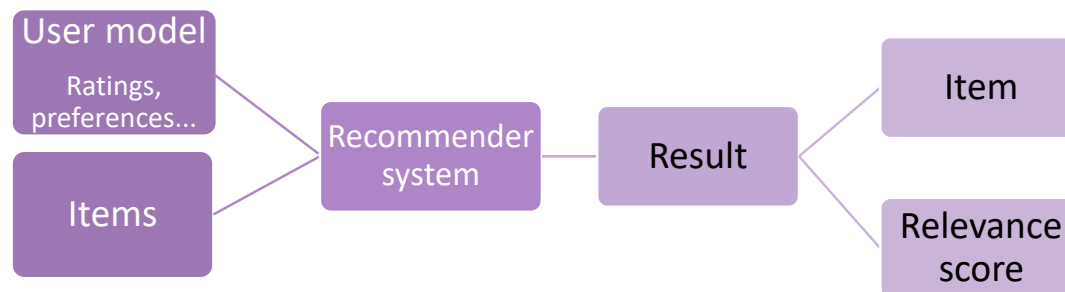
$$AP = \left(\frac{1}{1} \frac{1}{15}\right) + \left(\frac{2}{2} \frac{1}{15}\right) + \left(\frac{3}{3} \frac{1}{15}\right) + \left(\frac{4}{4} \frac{1}{15}\right) + \left(\frac{4}{5} 0\right) + \left(\frac{5}{6} \frac{1}{15}\right) + \dots$$

$$\text{Result} = [\quad 1, \quad \quad 1, \quad \quad 1, \quad \quad 1, \quad \quad 0, \quad \quad 1, \quad \dots]$$

Els punts en que la *recall* no canvia (els no-rellevants) no contribueixen a la suma.

SISTEMA RECOMANADOR

Ens trobem que no tenim *query*, ni y_{train} . És un model no supervisat. A més, redueixen la sobre càrrega d'informació, estimant la rellevància.



Filtratge col·laboratiu

Crea falsos positius, s'influencia per la tendència d'altres persones. I assumeix que els gustos dels usuaris no varien.

USER TO USER

Es mira quin usuari s'assembla més. Quan hi ha més usuaris que ítems és millor utilitzar Item-item, ja que proporcionarà millor prediccions.

ITEM TO ITEM

Relaciona els articles comprats o valorats per un usuari objectiu amb articles similars i combina aquests articles similars en una llista de recomanacions. La similitud basada en ítem és més estable que la d'usuari.

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

2023-24

Similitud

user - user

La correlació de Pearson és una mesura de similitud popular en el filtratge col·laboratiu basat en l'usuari. Retorna valors de semblança en l'interval de $[-1, 1]$.

$$\text{sim}(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

Values centred to average rating given by each user

a, b : users

$r_{a,p}$: rating of user a for item p

P : set of items, rated both by a and b

Values normalised by the variance of the ratings given by each user

També es fa servir la similitud de cosinus, de manera que es mira que els angles entre les rectes d'usuari sigui petit.

$$\vec{a} \cdot \vec{b} = |\vec{a}| * |\vec{b}| \cos(\angle \vec{a} \vec{b}) \quad \text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

WUOLAH

Item-item

Es sol fer servir la distància ajustada de cosinus.

$$sim_{AdjustedCosine}(i, j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_u)^2}}$$

Compare to cosine distance
 $sim_{cosine}(i, j) = \frac{i \cdot j}{|i| \cdot |j|}$

Prediccions

user-user

Una funció de predicció comuna és utilitzar les similituds per calcular una ponderació mitjana de classificacions (centrades).

$$pred(a, i) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,i} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

item-item

Fem servir una mitjana ponderada.

$$pred(u, p) = \frac{\sum_{i \in ratedItem(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ratedItem(u)} sim(i, p)}$$

FACTORITZACIÓ DE MATRIUS

Es sol fer servir quan treballem recomanacions basades en el contingut.

Amb la factorització de matrius s'intenta representar espai latent, variables que no estan indicades però que es tenen en compte, per exemple bag of words.

Per tant, busquem el millor espai latent (multiplicació de dues matrius) per entendre/explicar els ratings (tercera matriu).

SVD

SVD és una manera matemàtica de trobar aquests dos matrius més petites que minimitzen l'error d'aproximació resultant, la mitjana error quadrat (MSE). Minimitzem l'error mitjançant el Descens del Gradient.

Podem utilitzar les matrius resultants U i M per predir les qualificacions del conjunt de proves.

$$p_{ij} = \sum_{k=1}^K u_{ki} \times m_{kj}$$

On,

- m_{kj} és la matriu subjacent de l'espai subjacent
- u_{ki} és la matriu subjacent de matrius

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

perdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin
me lo quito yo...

WUOLAH

AC

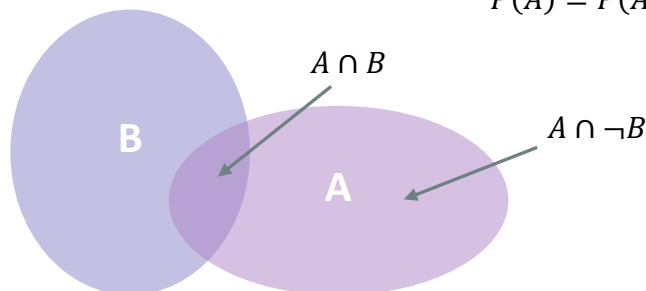
2023-24

PROBABILITATS

PROBABILITAT CONJUNTA

La probabilitat conjunta és la probabilitat de casos simultanis.

$$P(A) = P(A \cap B) + P(A \cap \neg B)$$



Exemple: Quanta gent amb colesterol dolent a 200 té infarts?

$$P(\text{tenir infart}) = P(\text{infart, colesterol} = 200) + P(\text{infart, colesterol} \neq 200) + \dots$$

PROBABILITAT CONDICIONADA

Fracció dels mons en els que X és cert donat que Y és cert,

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

WUOLAH

REGLA DE BAYES

A partir de la fórmula de la probabilidad condicionada, podemos decir que:

- $P(X|Y)$ representa el training set
- $P(Y)$, probabilidad de cada clase (ex. Infart o no)

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Exemple

Quina és la probabilitat que tinguis grip si acabes de tossir?

Assumim: $P(\text{grip}) = 0.05$ $P(\text{tossir} \mid \text{grip}) = 0.80$ $P(\text{tossir} \mid \sim \text{grip}) = 0.1$

Llavors,

$$P(T+) = P(T+, G+) + P(T+, G-)$$

$$P(T+, G+) = P(T+ \mid G+)P(G+)$$

$$P(T+, G-) = P(T+ \mid G-)P(G-)$$

$$P(G+ \mid T+) = P(G+ \mid T+)/P(T+) = P(T+ \mid G+)P(G+)/P(T+)$$

JOINT DISTRIBUTION

1. Fer una taula de veritat amb totes les combinacions dels valors de les variables (si hi ha M variables booleanes llavors tindrem 2^M files).
2. Per a cada combinació calcular com és de probable.
3. Aquestes probabilitats han de sumar 1 segons els axiomes de la probabilitat.

Exemple

A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10

$$P(C=1 | A=0) = P(C=1, A=0) / P(A=0) =$$

$$= P(C=1, A=0, B=0) + P(C=1, A=0, B=1)$$

$$P(A=0) = P(A=0, C=1) + P(A=0, C=0)$$

Importante

Puedo eliminar la publi de este documento con 1 coin ¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo espacio



Necesito concentración
ali ali ooh
esto con 1 coin
me lo quito yo...



WUOLAH

AC

2023-24

NAIVES BAYES

- Les característiques de Naives Bayes són:
- Aprenentatge i test molt ràpid (bàsicament comptar dades)
- Requeriments de memòria baixos.
- Molt bo en dominis amb moltes característiques igualment importants.
- Més robust a dades irrelevantes que molts altres mètodes.
- Més robust a 'concept drift' (canvi de definició de la classe sobre el temps).

SUPOSICIÓ NAÏVE

Existeix independència condicional d'atributs $P(x_1, \dots, x_k | C) = P(x_1 | C) \cdot \dots \cdot P(x_k | C)$,

Per evitar fer multiplicacions i tenir números massa petits, utilitzem logaritmes.

$$c = \arg \max_{C_j \in C} \left(\log P(C_j) + \sum_{a_i \in x} \log P(a_i | C_j) \right)$$

La hipòtesi d'independència condicionada fa possible la computació i dona un classificador òptim si es compleix, però rarament es satisfà a la pràctica, ja que els atributs (variables) sovint estan correlacionats. Per evitar aquesta limitació:

WUOLAH

Bayesian networks

Combinen el raonament bayesià amb relacions de causalitat entre atributs.

Decision trees

Raonen sobre un atribut a cada pas, considerant els atributs més 'importantes' primer.