
Gestió de característiques

PID_00284564

Jordi Gironés Roig

Temps mínim de dedicació recomanat: 1 hora



Universitat
Oberta
de Catalunya

Jordi Gironés Roig

Llicenciat en Matemàtiques per la Universitat Autònoma de Barcelona i diplomat en Empresarials per la Universitat Oberta de Catalunya. Ha desenvolupat gran part de la seva carrera professional al voltant de la solució SAP, en els seus vessants operatius amb S4-HANA i estratègica amb SAP-BI. Actualment treballa en la indústria químic-farmacèutica com a responsable d'aplicacions corporatives per a Esteve Pharmaceuticals i col·labora amb la UOC en assignatures relacionades amb l'anàlisi de dades.

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats pel professor: Julià Minguillón Alfonso

Primera edició: setembre 2021

© d'aquesta edició, Fundació Universitat Oberta de Catalunya (FUOC)

Av. Tibidabo, 39-43, 08035 Barcelona

Autoria: Jordi Gironés Roig

Producció: FUOC

Tots els drets reservats

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit del titular dels drets.

Índex

Introducció	5
1. Visualització de característiques	7
2. Selecció de característiques	10
2.1. <i>Undersampling</i>	11
2.1.1. SMOTE	11
2.1.2. ADASYN	11
2.2. Percentatge de valors nuls	11
2.3. Variància estadística	12
2.4. Correlació entre variables explicatives	12
2.5. PCA i SVD	13
2.6. Mesures d'importància	13
2.7. Processos seqüencials	14
2.8. <i>Stability selection</i>	15
3. Enginyeria de característiques	16
3.1. <i>Feature bucketing</i>	16
3.2. <i>Feature hashing</i>	17
3.3. <i>Feature crossing</i>	18
3.4. Enginyeria sobre dades categòriques	18
3.5. Tipus de codificació	19
3.6. Arrodoniment	19
3.7. Transformació logarítmica	19

Introducció

La meitat de la capacitat del cervell humà és responsable de processar menys del 5 % de l'espectre visual al qual tenim accés. Ni tan sols el nostre sistema nerviós òptic té l'habilitat de capturar totes les dades que té al seu abast.

I, no obstant això, la nostra capacitat visual és reconegudament útil, eficaç i eficient. En aquest escenari, sembla lògic plantejar-se la pregunta següent: com s'aconsegueixen tals nivells d'excel·lència amb uns recursos tan limitats?

Simplificació és la paraula clau. **Prescindir del que no serveix**, focalitzar-se en allò rellevant, identificar patrons i aprendre. Aquests són els pilars en els quals es basa el nostre cervell per operar enmig de tant de caos i aquest serà el nostre *leitmotiv* per entendre bé la gestió de característiques en la mineria de dades.

Un dels nostres genis de la pintura universal, Pablo Picasso, descrivia d'aquesta manera el procés creatiu:

«L'art és l'eliminació d'allò innecessari».

Aquesta frase conclou la millor descripció que es pot fer d'un procés com la **selecció de característiques**. Tant és així que treballar amb atributs no adequats pot repercutir negativament en la construcció de models analítics de les maneres següents:

- Alentint el procés d'aprenentatge.
- Dificultant la interpretabilitat del model generat.
- Disminuint dràsticament la capacitat de generalització del model.

Podem anotar una nova frase del mestre pintor per il·lustrar un altre aspecte important de la gestió de característiques:

«Alguns pintors transformen el sol en una taca groga, uns altres transformen una taca groga en un sol».

Transformar un atribut en brut en un atribut capaç de millorar la capacitat predictiva d'un model és el que coneixem com a *enginyeria de característiques* i convertirà el nostre joc de dades inicial en un joc de dades que definitivament descriu millor el problema analític que provem d'estudiar.

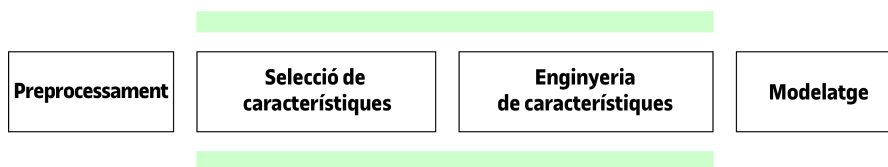
Quan un model passa de consumir aliments crus a aliments elaborats i més ben adaptats a les seves necessitats nutricionals rendeix millor en tots els seus aspectes.

D'aquesta manera, l'enginyeria de característiques aportarà als nostres projectes analítics els beneficis següents:

- L'increment de la capacitat d'aprenentatge.
- La simplificació de la interpretabilitat del model generat.

En la figura 1 posem en context dins del procés analític les fases de selecció i enginyeria de característiques.

Figura 1. Gestió de característiques



A tall de resum, una bona gestió de característiques, basada en la **selecció d'atributs** realment rellevants per al nostre problema analític i en una **enginyeria d'atributs** orientada a optimitzar el seu posterior ús en models, serà el que en moltes ocasions marcarà la diferència. En aquest material docent introduïm breument diferents mètodes per a la gestió de característiques, sense entrar en detalls tècnics.

1. Visualització de característiques

Un dels consells més repetits entre els experts en intel·ligència artificial és aquest: visualitza les teves dades amb certa freqüència.

La visualització de dades tradicionalment se situa en la fase de preparació del joc de dades, però és també recomanable usar les capacitats gràfiques en fases de transformació de característiques i modelatge.

Però què busquem en les visualitzacions? I quins beneficis poden aportar-nos? Bàsicament són dos:

- 1) Ens ajudaran a **corroborar les nostres suposicions** en cada pas que donem.
- 2) Ens permetran **valorar els efectes dels canvis** que anem introduint en el joc de dades.

Els tipus de gràfics més utilitzats són els següents:

- Gràfics de punts, en els quals podem visualitzar la relació entre dues variables.
- Gràfics de barres, que s'usen per visualitzar dades categorials en barres verticals proporcionals als valors que representen.
- Histogrames de freqüències, que ens permeten analitzar com de comuna és una classe en una característica determinada.

Hi ha molts més tipus de gràfics, com els de caixes, els de sectors o els de densitat, i haurem de valorar quin s'adapta millor a les necessitats de visualització de les dades que anem detectant.

Gestionar el format del gràfic també pot contribuir a facilitar la interpretació i la comprensió del seu contingut. Alguns aspectes que cal considerar són:

- Assignar colors i donar-los un significat d'acord amb les dades.
- Usar altres figures geomètriques, a més de punts, i assignar, també, un significat a cada figura.

- Elaborar les llegendes, explicant el significat de cada símbol.
- Acompanyar el gràfic amb la taula de valors que el genera, ja que a vegades pot ajudar.

Per mostrar com les visualitzacions ens permetran anar més enllà de les mesures estadístiques clàssiques, prendrem el joc de dades conegut com a *quartet d'Anscombe* que ens mostra la figura 2.

Figura 2. Joc de dades del quartet d'Anscombe

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

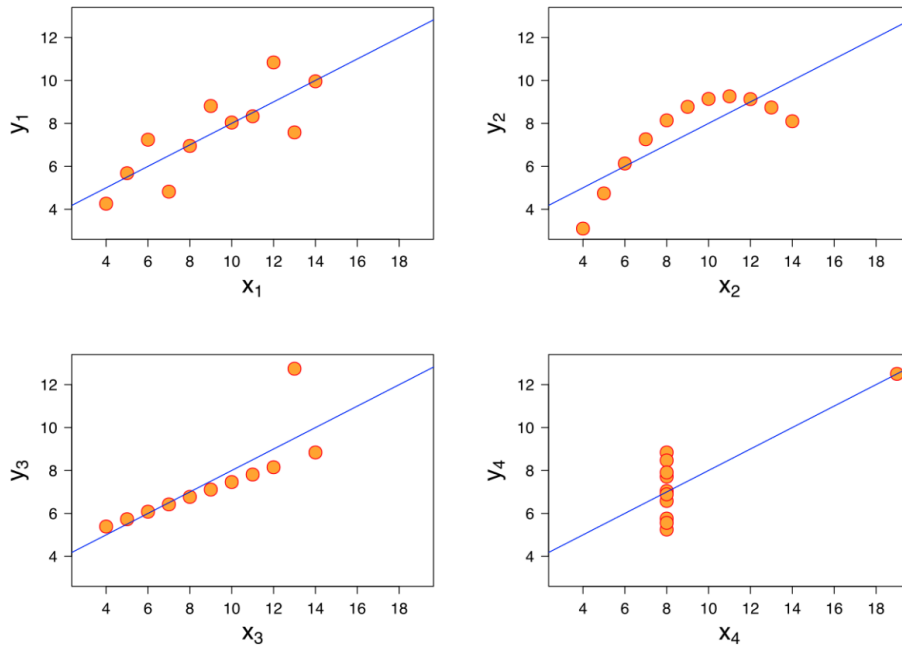
Els quatre jocs de dades comparteixen les mesures estadístiques següents:

- Mitjana (x) = 9
- Variància (x) = 11
- Mitjana (y) = 7.50
- Variància (y) = 4.125
- Correlació entre x i y = 0.816
- Recta de regressió $y = 3 + 0.5 x$

En la figura 3 podem veure com, malgrat que els quatre jocs de dades comparteixen les mateixes estadístiques, tenim visualitzacions molt diferents:

- Cas I: veiem una relació lineal entre les variables x i y .
- Cas II: amb relació parabòlica.
- Cas III: amb relació lineal, però amb un valor extrem significatiu.
- Cas IV: un sol valor extrem és capaç de produir un coeficient de correlació alt malgrat que la resta de punts no presentin cap correlació.

Figura 3. Visualitzacions sobre el quartet d'Anscombe



En definitiva, les visualitzacions ens permetran aquest pas enrere que proporciona la perspectiva necessària per veure el que *a priori* no és obvi, però hi és.

J. A. Baker, en el seu llibre *The Peregrin*, escrivia:

«The hardest thing of all to see is what is really there».

A vegades, el més difícil de veure és precisament el que tenim davant nostre i és per aquest motiu que, malgrat que sigui una eina irrenunciable, les visualitzacions no són suficients per captar el coneixement subjacent en un joc de dades.

2. Selecció de característiques

Comencem per aclarir què entenem per característica. Tal com es pot veure en la figura 4, en un joc de dades trobarem files i columnes, en què les primeres representen observacions o esdeveniments descrits pel joc de dades i les segones representen les propietats d'aquestes observacions.

Figura 4. Definició de *característica*

	Article	Unitats	Vendes	← Característiques
Observacions →	Catifa	500	450	
Observacions →	Cadira	800	950	
Observacions →	Prestatgeria	200	300	

Si ho traslладem al món del llenguatge, podem pensar en les files com si fossin substantius i en les columnes com si fossin adjectius que defineixen i qualifiquen els substantius. Al llarg d'aquest material didàctic ens referirem a les columnes també com a *atributs*, *característiques* i *variables*.

La idea generalitzada que com més dades millor és del tot errònia i hauria de ser substituïda per aquesta altra: com més dades rellevants per al nostre objectiu, millor.

Sota el concepte *feature selection*, en la seva accepció anglesa, englobem una sèrie de tècniques orientades a seleccionar només aquelles característiques en un joc de dades que ens aportaran i faran que el joc de dades realment descrigui el problema analític que tractem de resoldre.

A continuació veurem com algunes d'aquestes tècniques són extremadament simples mentre que d'altres són més elaborades, malgrat tot, no hem d'oblidar que totes persegueixen un objectiu comú:

Aconseguir que el joc de dades representi tan bé com sigui possible el problema analític que tractem de resoldre.

2.1. Undersampling

La circumstància de l'*oversampling* i de l'*undersampling* impacta especialment en els problemes de classificació i succeeix quan una de les classes objectiu està sobrerrepresentada o infrarepresentada en el joc de dades.

Això pot provocar que l'algorisme de classificació només aprengui de la classe de la qual té més informació i, en conseqüència, generi un model que simplement ignori la classe amb menys representació en el joc de dades.

Per evitar aquest tipus de situacions disposem de les tècniques SMOTE i ADASYN.

2.1.1. SMOTE

Acrònim del terme anglès *synthetic minority oversampling technique*. És una tècnica estadística que permet generar en un joc de dades observacions noves que ajudin a incrementar la representativitat d'aquelles classes minoritàries en el joc de dades.

En lloc de generar observacions noves simplement com a còpia de les ja existents, el que fa SMOTE és prendre els veïns més pròxims (K-NN) de les observacions amb classe minoritària i generar una combinació de les seves característiques.

2.1.2. ADASYN

Acrònim del terme anglès *adaptive synthetic sampling approach*. Es tracta d'una evolució de l'anterior algorisme SMOTE, de manera que en lloc de tractar de replicar les observacions amb classes minoritàries, només tracta de fer-ho sobre aquelles classes minoritàries amb dificultat d'aprenentatge alt.

2.2. Percentatge de valors nuls

Els valors nuls, o *missing values* en anglès, poden ocasionar problemes en el funcionament de certs algorismes, per la qual cosa és una situació que cal gestionar des de diferents aproximacions. Les més habituals són:

- Substituir-los per la mitjana estadística.
- Substituir-los per la mediana estadística.
- Substituir-los per valors aleatoris.
- Substituir-los per zero.
- Eliminar les files afectades.

Des del punt de vista de la gestió de característiques, hauríem de plantejar-nos una possibilitat més: eliminar les característiques que superin un percentatge preestablert de valors nuls. Més concretament, una possible estratègia seria la següent:

- Per sobre del 95 % de valors nuls es descarta la característica.
- Entre el 50 % i el 95 % de valors nuls, aquests no s'emplenen.
- Per sota del 50 % de valors nuls, aquests s'emplenen seguint alguna de les opcions ja esmentades.

Aquesta estratègia, basada en el grau d'afectació dels valors nuls sobre cada característica, ens permetrà una gestió més detallada del problema.

2.3. Variància estadística

En un joc de dades, la quantitat d'informació continguda en una característica X amb valors x_1, \dots, x_n , es pot mesurar mitjançant la variància estadística:

$$\text{VAR}(X) = \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (1)$$

On x_i són els valors de la característica X i μ és la seva mitjana estadística.

Cal descartar les característiques amb una variància pròxima a zero, ja que, en manca de variabilitat, no contenen informació i, en conseqüència, no descriuen cap propietat de les entrades del joc de dades.

Mantenir característiques amb variància molt baixa només contribuirà a complicar el model i a confondre'l.

2.4. Correlació entre variables explicatives

Si dues característiques explicatives estan molt correlacionades entre elles significa que ambdues aporten informació redundant al joc de dades i, en conseqüència, podem prescindir d'una d'elles.

Amb quina de les dues característiques ens quedarem? Amb aquella que tingui més correlació amb la variable objectiu, *target* en anglès.

La situació ideal seria que totes les variables explicatives fossin independents entre elles, és a dir, que hi hagués correlació zero entre elles i només la tinguessin amb la variable objectiu.

No obstant això, cal tenir molta cura quan treballem amb correlacions perquè aquestes mesuren la relació lineal entre característiques i no hem d'oblidar que **aquesta relació podria existir i no ser lineal**.

Una conseqüència de la frase anterior és que, si una variable explicativa no està correlacionada amb la variable objectiu o *target*, no hem d'arribar a la conclusió que podem prescindir de la variable explicativa, ja que així i tot, aquesta pot contenir informació rellevant per predir la variable objectiu.

2.5. PCA i SVD

Tant l'anàlisi de components principals, *principal component analysis* (PCA) en anglès, com la descomposició de valors singulars, *singular value decomposition* (SVD) en anglès, són tècniques que ens permeten treballar amb noves característiques anomenades components, que certament són independents entre si.

En realitat, aquestes dues tècniques ens permeten representar el joc de dades en un nou sistema de coordenades que anomenem *components principals*. Aquest sistema està més ben adaptat a la distribució del joc de dades, de manera que recull millor la seva variabilitat.

El que se sol fer és **incorporar al joc de dades original**, *raw data set* en anglès, els **components principals** com a característiques noves, seleccionant aquells components que, de manera acumulativa, recullen una variabilitat superior al 90 %. Aquesta acció pot actuar de catalitzador en un procés de classificació.

Com a contrapartida, tenim l'inconvenient de la interpretabilitat, és a dir, donar significat a aquestes noves dimensions en el joc de dades.

2.6. Mesures d'importància

Els arbres de decisió són algorismes de classificació que construeixen camins de decisió a partir de les variables del joc de dades, les quals participen per ordre d'importància, de manera que primer intervenen les característiques més rellevants per al procés de classificació.

Això significa que són capaços d'ordenar les característiques segons la seva importància o rellevància en el joc de dades. Però com podem mesurar la importància d'un atribut en un joc de dades?

Una de les formes més habituals de mesurar aquesta importància és el **guany de la informació** o petites variacions sobre aquest concepte.

Al seu torn, el guany d'informació recau sobre un altre concepte: l'**entropia**.

$$E(X) = \sum_{i=1}^n -p_i \log_2 p_i \quad (2)$$

En aquest concepte, p_i és la probabilitat que es doni el valor x_i en l'atribut X . D'aquesta manera, podem dir que l'entropia mesura el desordre o la quantitat d'informació continguda en un joc de dades.

A partir d'aquí, si som capaços de mesurar la variació que es produeix en l'entropia d'un joc de dades cada vegada que fixem un valor d'un atribut, tindrem un criteri per decidir quin atribut maximitza el guany d'informació, *information gain* en anglès.

$$IG(T) = H(T) - H(T|x) \quad (3)$$

Aquí, T és un joc de dades amb diversos atributs i $H(T|x)$ és l'entropia del joc de dades T quan fixem un valor de l'atribut x .

Tant utilitzar directament el guany d'informació com utilitzar els diferents algorismes d'arbres de decisió són bones opcions per ordenar les característiques d'un joc de dades tenint en compte la seva importància i seleccionar només aquelles que són més rellevants.

2.7. Processos seqüencials

Podem usar processos seqüencials d'addició o sostracció de característiques a un joc de dades si s'ha de jutjar per la seva rellevància amb l'objectiu de maximitzar algun criteri preestablert.

Aquest criteri podria ser el nivell de precisió d'un algorisme predictiu. D'aquesta manera, aplicariem processos seqüencials de selecció de característiques sobre el joc de dades, fins que el model que consumeix les diferents versions de jocs de dades que anem generant aconsegueixi un nivell de rendiment acceptable.

Bàsicament, podem distingir tres tipus de processos seqüencials de selecció de característiques:

1) *Forward selection*

- Identificar les característiques més rellevants.
- Afegir la característica següent al model.
- Iterar el pas anterior fins a aconseguir un nivell de precisió satisfactori en el model.

2) *Backward elimination*

- Executar el model considerant totes les característiques.
- Eliminar la característica menys rellevant i tornar a aplicar el model.
- Iterar el pas anterior fins a aconseguir un nivell de precisió satisfactori en el model.

3) *Stepwise selection*. Aquest procés és igual que el de *forward selection*, excepte pel fet que, al llarg de les iteracions, ens plantejarem si val la pena eliminar alguna característica que hagi esdevingut irrellevant.

2.8. *Stability selection*

La selecció de característiques per estabilitat consisteix en aplicar un algorisme de selecció de característiques de forma repetida sobre diferents subconjunts o mostres del joc de dades original. La mostra es pren tant a nivell d'observacions com a nivell de característiques.

Després de repetir el procés anterior un cert nombre de vegades, agreguem els resultats de manera que puguem verificar quantes vegades s'ha considerat important una característica.

D'aquesta manera podrem distingir tres tipus de característiques:

- Característiques molt rellevants, que identificarem per tenir puntuacions pròximes al 100 %.
- Característiques més febles, però igualment rellevants, que identificarem per tenir puntuacions per sobre de 0.
- Característiques irrellevants, que identificarem per tenir puntuacions pròximes a 0.

En l'article «Stability selection, RFE and everything side by side» es pot trobar un bon exemple d'aquest procés.

Referència bibliogràfica

«Stability selection, RFE and everything side by side» (2014). *Diving into data*. <<https://bit.ly/2rlmhz3>>

3. Enginyeria de característiques

L'èxit de molts algorismes de mineria de dades rau en com presentem les dades al model.

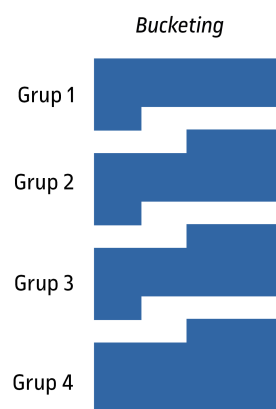
Un problema analític ben plantejat dependrà, principalment, de tres factors: disposar de les dades adequades, presentar les característiques ben estructurades i aplicar el model correcte.

Fins i tot si no seleccionem el millor dels models possibles, si les dades tenen potencial i les característiques es presenten adequadament, podrem obtenir bons resultats. D'aquesta manera, val la pena prendre's l'enginyeria de característiques com un **aspecte estratègic** en qualsevol estudi analític de dades.

3.1. *Feature bucketing*

L'agrupació de valors numèrics en rangs d'acord amb el problema analític que tractem de resoldre ajuda a millorar el rendiment en algorismes de classificació com les xarxes neuronals. En realitat, també pot veure's com una transformació de variables numèriques en categorials.

Figura 5. *Feature bucketing*



Quan treballem amb variables contínues com, per exemple, les vendes d'un producte, la quantitat de sucre per litre, els ingressos d'un client, etc., els al-

gorismes de classificació prenen cada valor com una classe individual, contribuint, així, al sobreentrenament i creant confusió sobre el que realment és important tenir en compte.

Podem considerar diferents criteris de categorització:

- Prendre la distribució per quartils.
- Prendre distribucions equidistants.

Agrupar per rangs els nostres valors continus contribuirà a la capacitat predictiva i descriptiva dels models analítics.

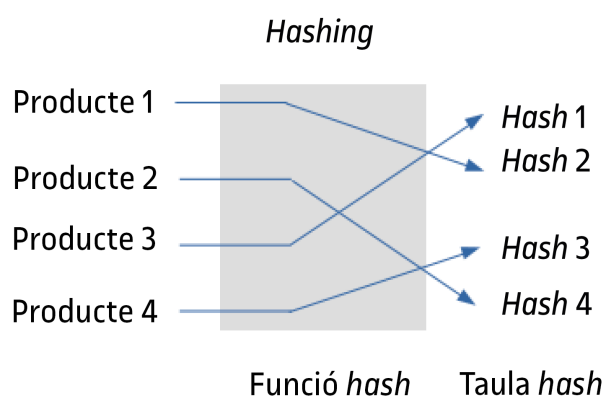
3.2. *Feature hashing*

Les funcions *hash* s'han convertit en una de les tècniques més usades dins del món de la intel·ligència artificial pels avantatges que ofereixen.

Hi ha molts tipus de funcions *hash*, però totes segueixen el mateix principi: mapen un grup de dades de grandària arbitrària contra un grup de dades de grandària fixa.

Prenent com a exemple la figura 6, si introduïm un hipotètic producte 5, la funció *hash* assignarà igualment un dels valors del diccionari {Hash 1, Hash 2, Hash 3, Hash 4}.

Figura 6. *Feature hashing*



D'aquesta manera, com que la característica resultant sempre tindrà el mateix diccionari de dades, no serà necessari entrenar de nou els models en els quals hagi participat en cas que la característica inicial sí que tingués valors nous.

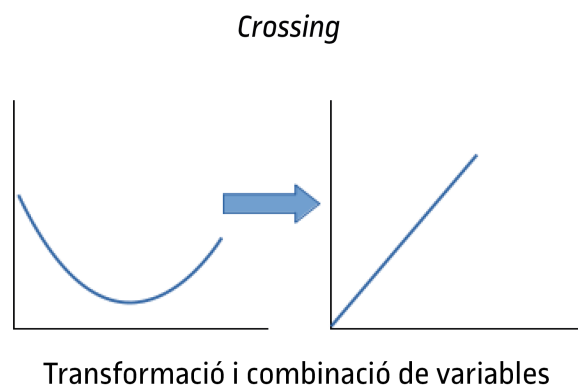
Precisament gràcies a aquesta habilitat, les tècniques de *feature hashing* estan especialment indicades per a variables categorials amb molts possibles valors, perquè d'alguna manera són capaces de comprimir-los.

Com a inconvenient, val la pena comentar que les funcions *hash* tenen el risc de generar col·lisions, és a dir, que valors inicials totalment dispars acabin sent assignats mitjançant el mapatge al mateix valor del diccionari, relacionant-los innecessàriament.

3.3. *Feature crossing*

L'encreuament de característiques, o *feature crossing* en anglès, és una tècnica senzilla, però que ens permetrà transformar problemes de classificació no lineals en problemes linealment classificables, tal com es pot veure a la figura 7.

Figura 7. *Feature crossing*



A tall d'exemple, en un joc de dades amb dues variables numèriques $\{x_1, x_2\}$ podrem considerar les següents tres noves variables $\{x_1^2, x_2^2, x_1 \cdot x_2\}$ amb l'objectiu que, després d'aquestes transformacions i considerant les noves variables, el nostre joc de dades passi a ser linealment separable.

3.4. Enginyeria sobre dades categòriques

Com que alguns models treballen millor amb atributs numèrics que amb atributs categorials, pot ser d'utilitat aplicar un procés de transformació a valors numèrics.

Aquest procés simplement consisteix en assignar un nombre decimal a cada classe de l'atribut original. Aquesta assignació té l'inconvenient que no té sentit mesurar distàncies o similituds sobre la variable decimal resultant, però podem mitigar aquest problema considerant també la codificació binària i la codificació *one-hot*.

3.5. Tipus de codificació

Prenent com a exemple la figura 8, podem veure que el diccionari de mitjans de transport urbà format per les classes {moto, bicicleta, patinet, monocicle, *segway*} el podem codificar de tres maneres diferents:

- **Decimal:** assignant un nombre decimal a cada classe.
- **Binari:** assignant un número expressat de manera binària.
- **One-hot:** assignant un vector de longitud al nombre de classes, amb 0 en totes les posicions excepte en la posició de la pròpia classe, que prendrà valor 1.

Figura 8. Tipus de codificació

Encoding

Etiqueta	Decimal	Binària	<i>One-hot</i>
moto	0	000	00001
bicicleta	1	001	00010
patinet	2	010	00100
monocicle	3	011	01000
<i>segway</i>	4	100	10000

Les codificacions binària i *one-hot* són habituals en algorismes com ara les xarxes neuronals.

3.6. Arrodoniment

A vegades, característiques numèriques com ara percentatges o proporcions solen expressar-se amb precisió decimal. Hem de preguntar-nos, en aquests casos, si realment necessitem aquest nivell de precisió per al nostre model i, en conseqüència, plantejar-nos l'opció d'arrodonir-les.

3.7. Transformació logarítmica

La transformació logarítmica ens ajudarà especialment en variables numèriques amb valors que poden prendre magnituds molt dispars.

Per exemple, si seleccionem una mostra d'habitatges d'una gran ciutat i considerem com a característica el seu valor de mercat, veurem molta discrepància

entre barris perifèrics i barris cèntrics, de manera que dificultarà la seva representació gràfica. A més, a vegades observarem també distribucions esbiaixades.

$$x \rightarrow \log_2 x \quad (4)$$

La transformació logarítmica a l'hora de considerar les potències de les magnituds contribueix a aconseguir una distribució més propera a la normal i a facilitar la seva representació gràfica, ja que aconsegueix magnituds més uniformes.

Per **reduir el biaix a la dreta** en una distribució de valors d'una variable usarem la **transformació logarítmica**.

Per **reduir el biaix a l'esquerra** en una distribució usarem la **transformació potencial**, que consistirà a prendre els quadrats, cubs i, fins i tot, potències superiors. La **transformació Box-Cox** també pertany a aquesta família i és cada vegada més usada en processos de *machine learning*.