Nearest Neighbor classification

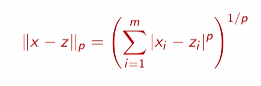
-Vectoritzar les dades d’entrenament amb etiquetes, que identifiquen els atributs de cada element.

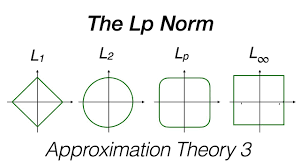
-Vectoritzar l’element a evaluar i calcular la distància mínima entre elli les dades d’entrenament. S’utilitza la funció de distancia euclidea per calcular la distància.

-S’extreu l’etiqueta de l’element que ha generat la menor distancia, obtenint els atributs que identifiquen l’element a evaluar.

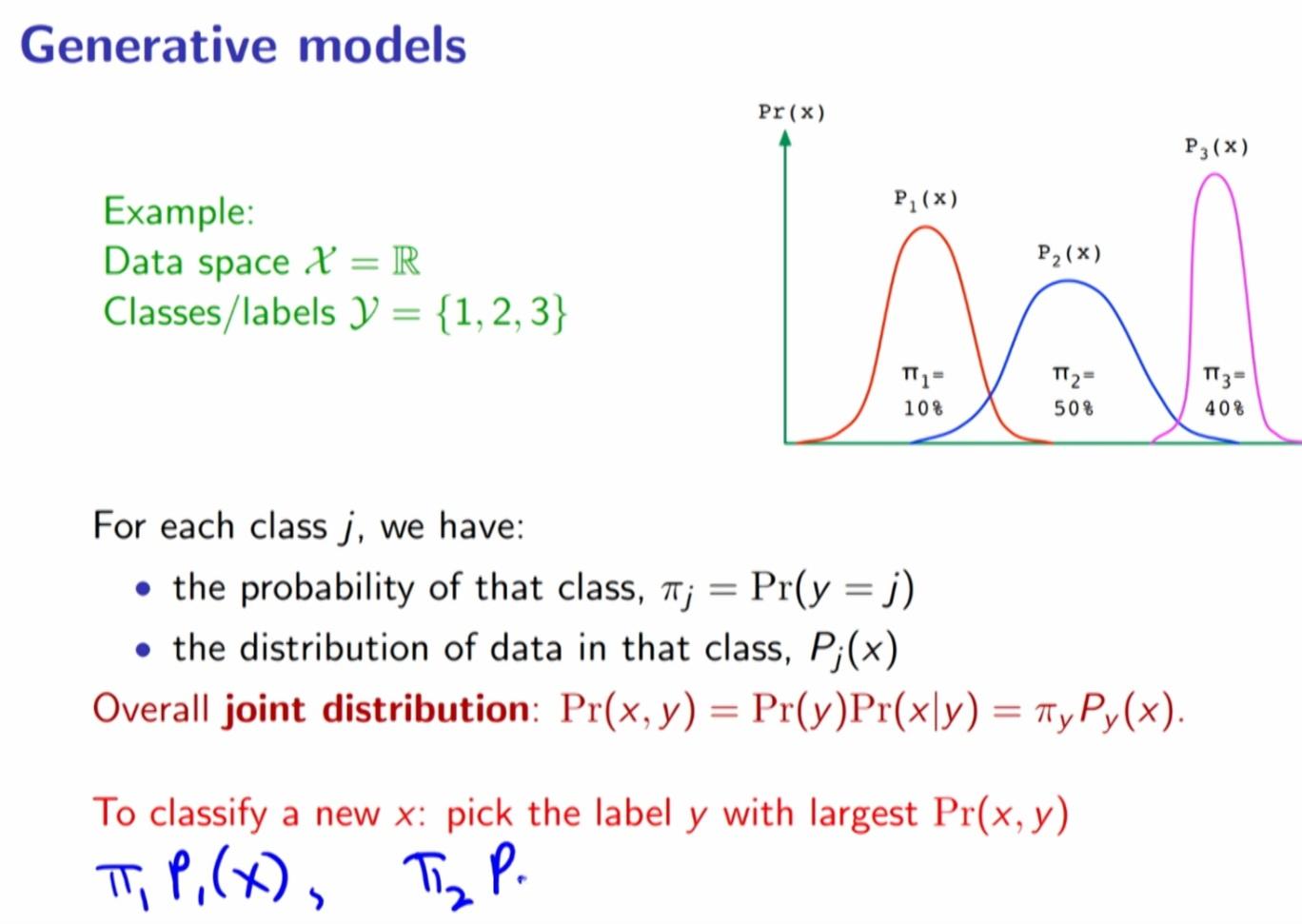
-Eliminar abans de l’entrenament totes les variables innecessàries, ja que alteren la distància.

-Funció de distància lp:





Models Generatius (basic)

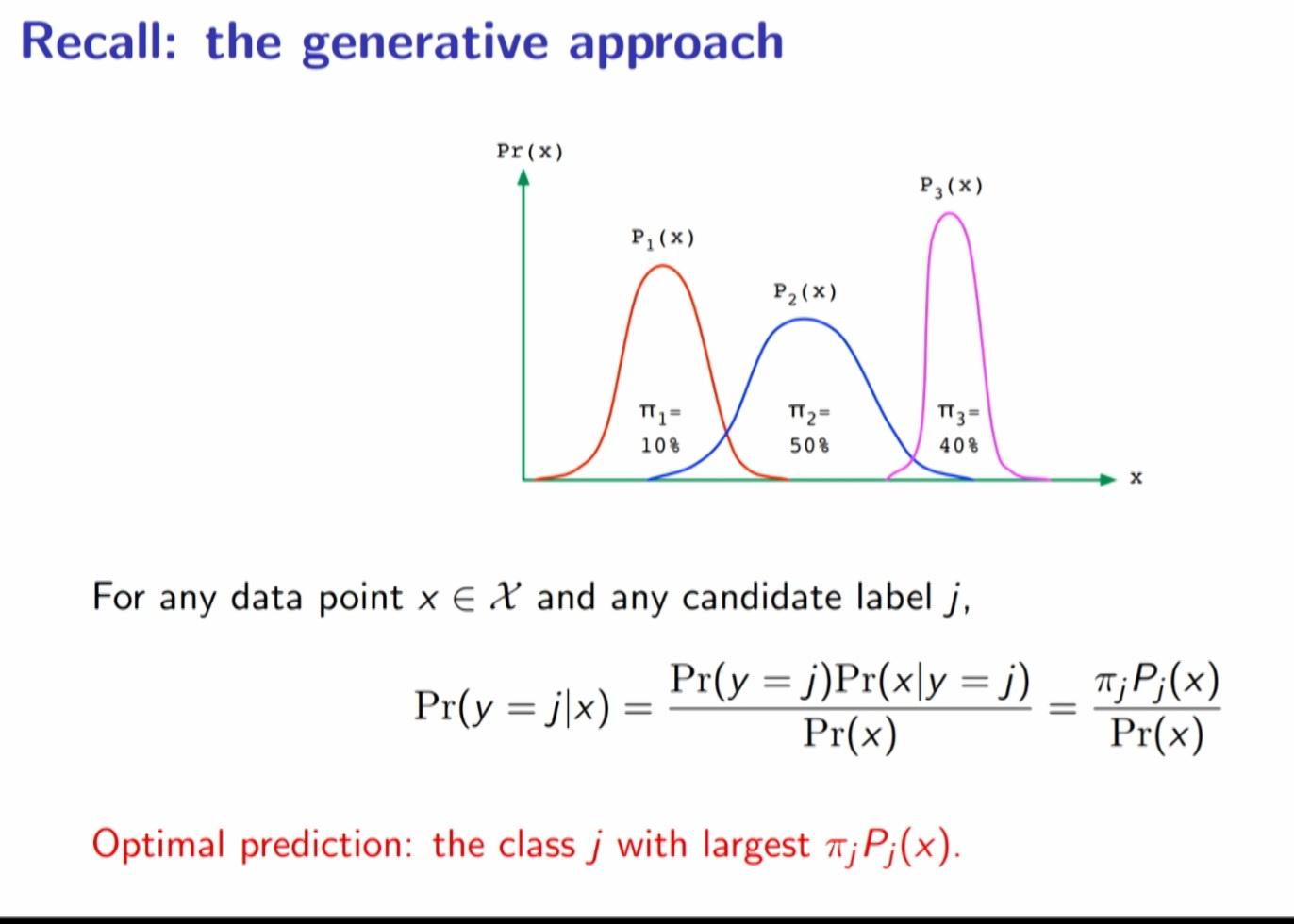


-A partir de les mostres obtingudes d’un experiment, es calcula la probabilitat a priori de cadascun dels resultats possibles (classes) mitjançant probabilitat bàsica (π\_i).

-A partir de paràmetres estadístics (variància, esperança, covariancia…), es modela una distribució de probabilitat per a cadascuna de les classes, que es independent de y.

-La distribució conjunta de x,y (P(x,y)) és el producte probabilitat a priori, distribució de y en funció de x.

-Per realitzar una predicció, s'escull el valor x sobre el qual predir i s’itera sobre tots els resultats (Y) possibles el càlcul de la probabilitat conjunta. El resultat major d’aquest càlcul és escollit i s’associa amb el resultat (y\_i) que l’ha generat. (y\_i) serà la predicció.

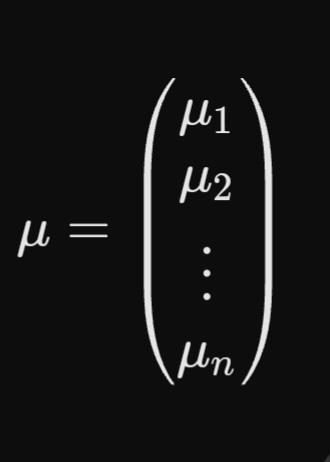


Formulació alternativa.

Gaussianes n-dimensionals

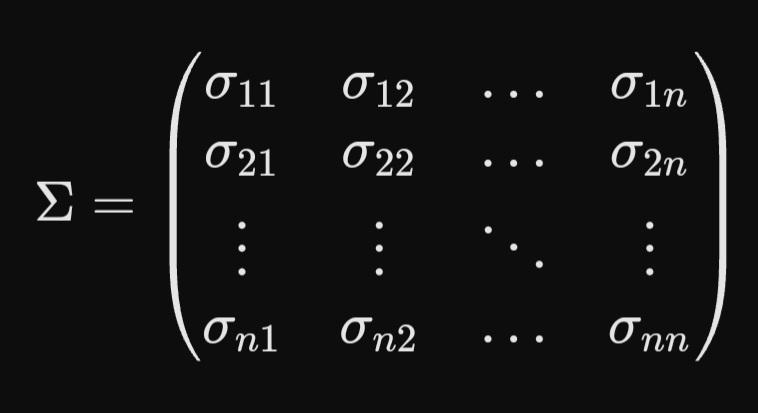
-En la gran majoria de casos, on les prediccions es realitzen sobre n-característiques, s’utilitzen gaussianes n-dimensionals com a distribució.

-Vector de mitjanes:



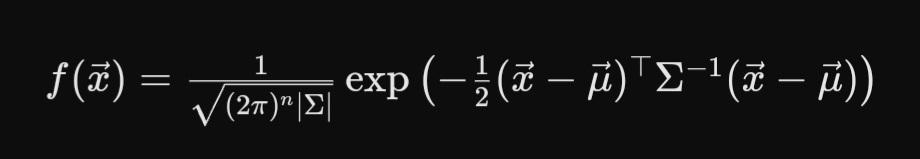
μ\_i representa la mitjana de cada gaussiana associada a cada variable aleatoria.

-Matriu de covariancies:



Es defineix com una matriu on es conté la informació de totes les variàncies i covariàncies de les n-variables de la gaussiana. Els elements diagonals (i=j) son les variàncies de les variables i, mentre la resta d'elements són les covariancies de les variables i i j.

-Es defineix la gaussiana n-dimensional mitjançant els paràmetres anteriors:

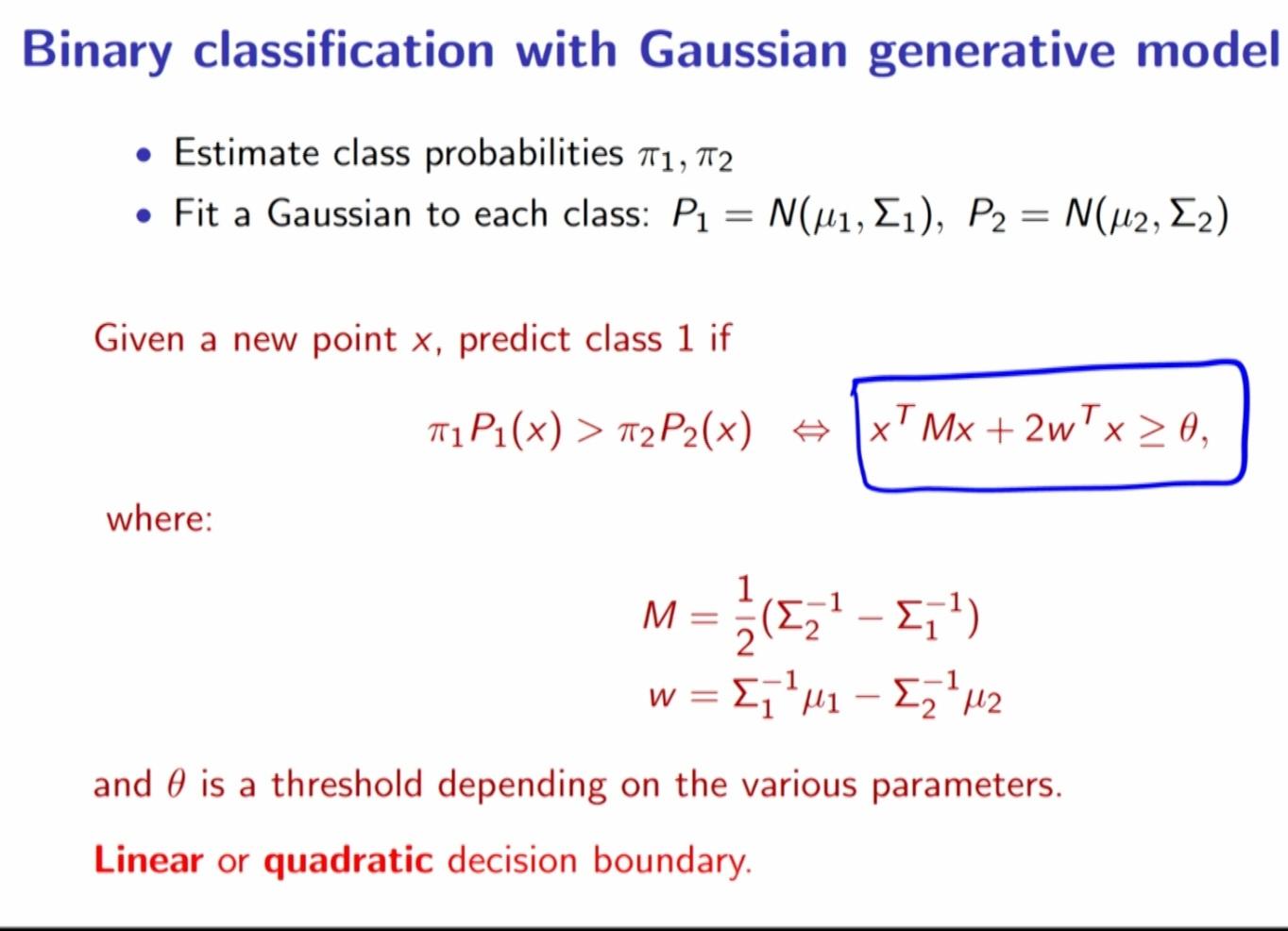


On s’utilitza el determinant i la inversa de la matriu de covariancies. El terme de l’exponencial representa la “distància quadràtica” entre **x** i **μ.**

**Un model gaussià n-Dimensional té en compte les correlacions entre variables.**

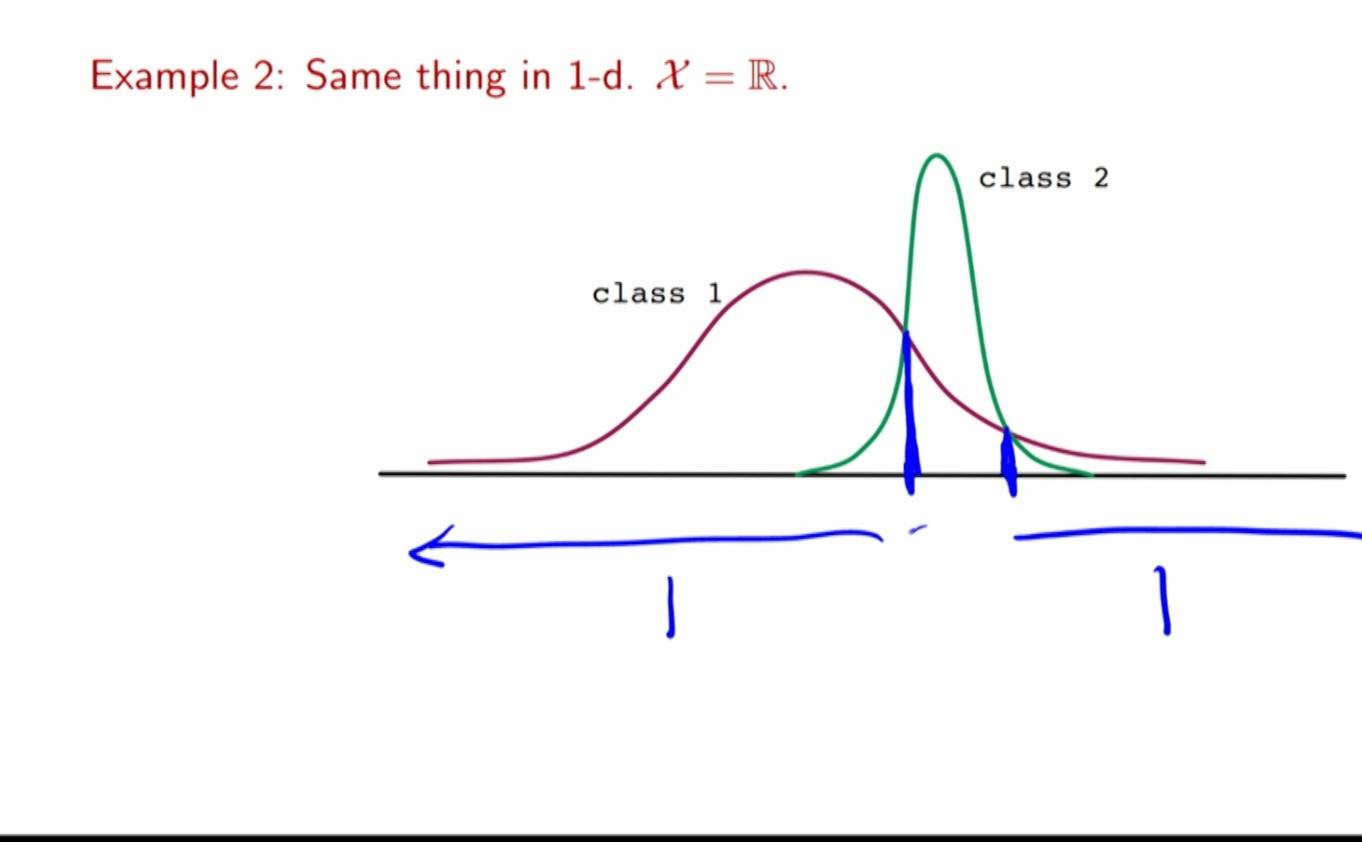
Classificació binària mitjançant un model generatiu gaussià

A part de realitzar prediccions com ja s’ha explicat anteriorment, resulta interessant conèixer la frontera de decisió d’un model generatiu. En el cas gaussià:



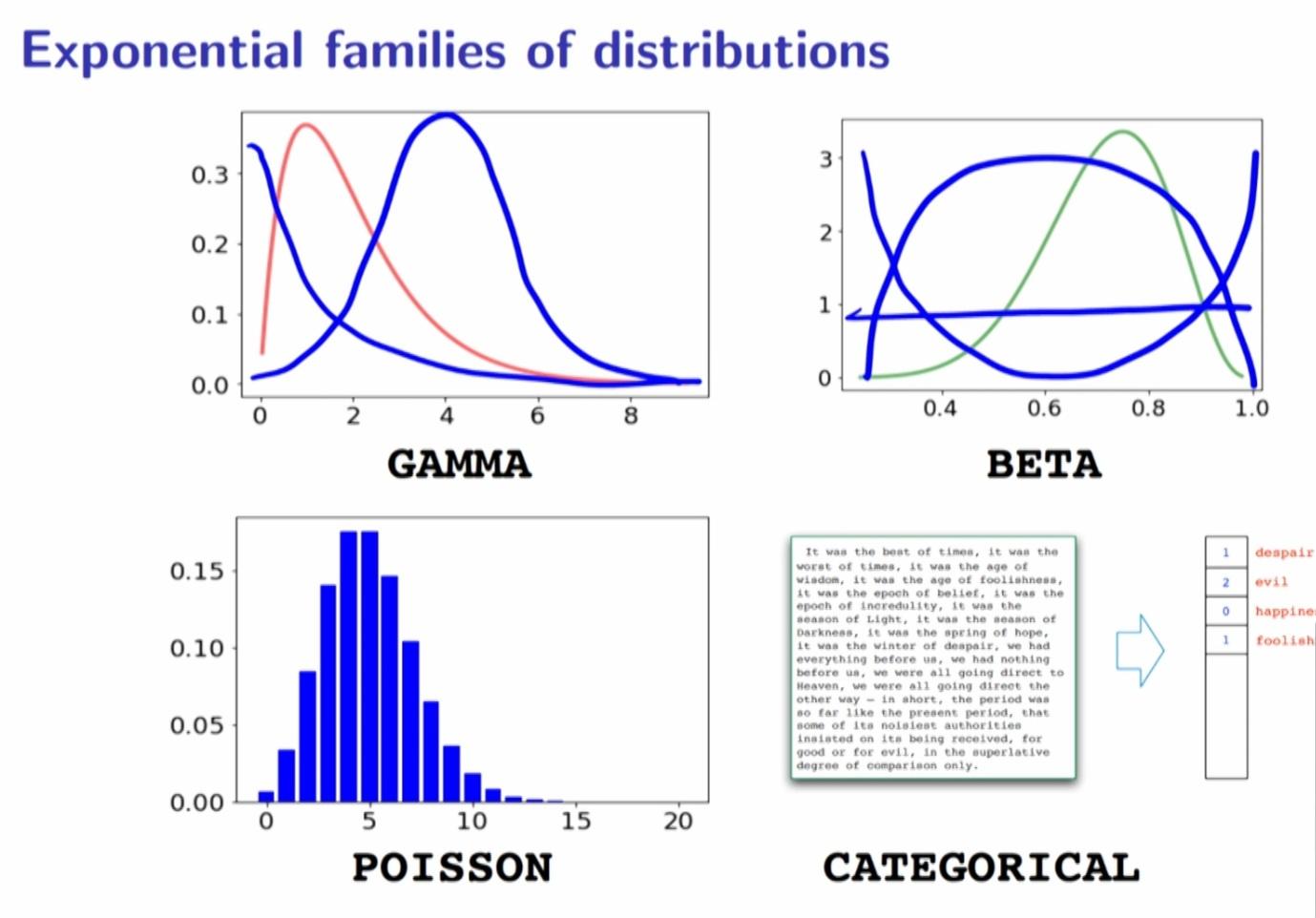
Per obtenir exactament la frontera, theta es planteja com una igualtat que traça una corba.

Exemple en R:

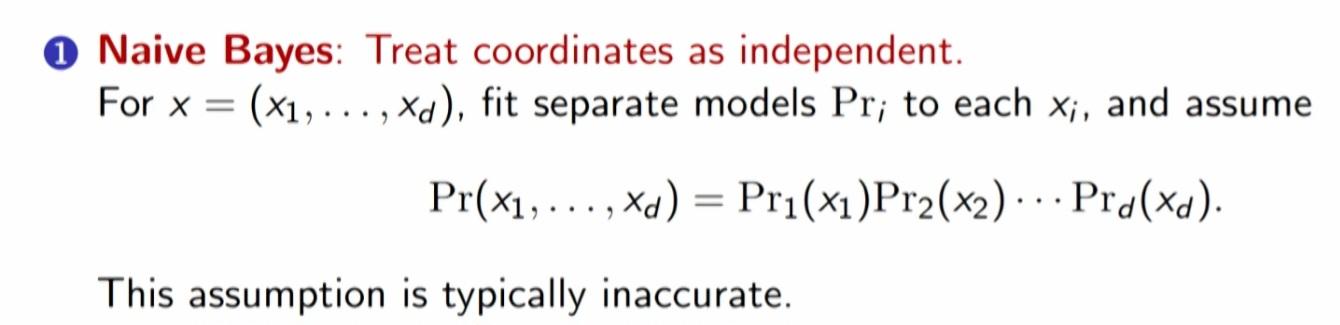


La frontera de decisió es traça on intersecten les dues gaussianes.

Distribucions exponencials

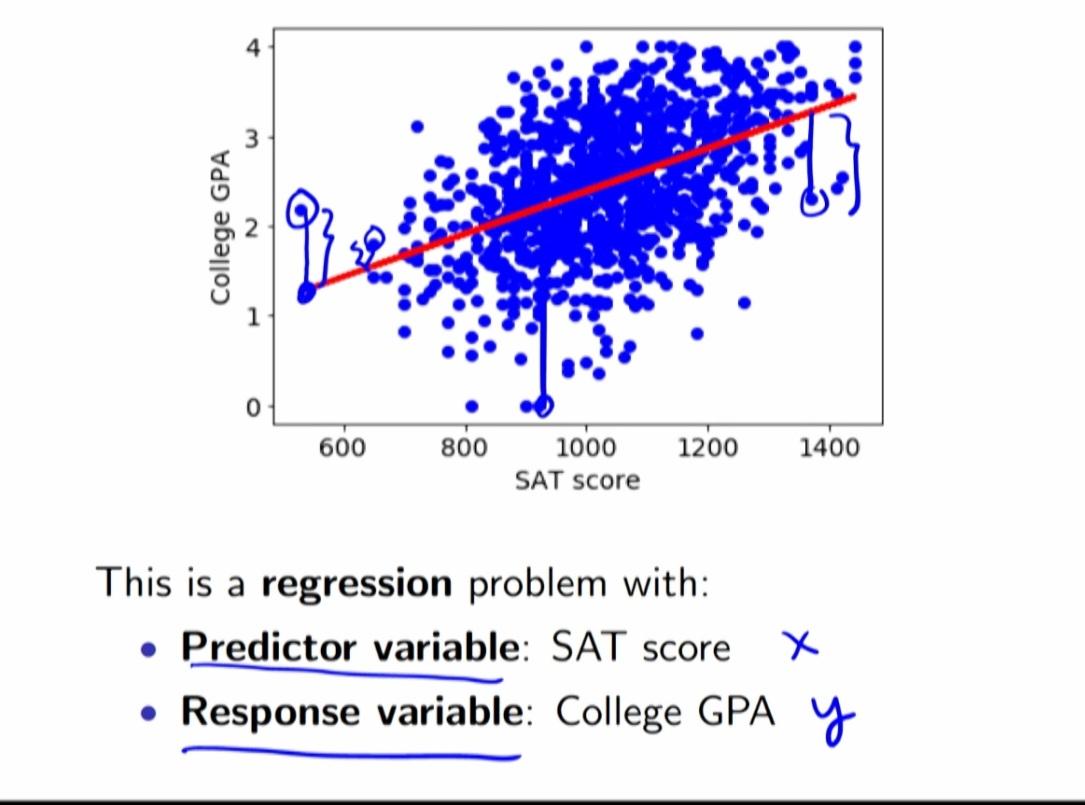


Naive Bayes

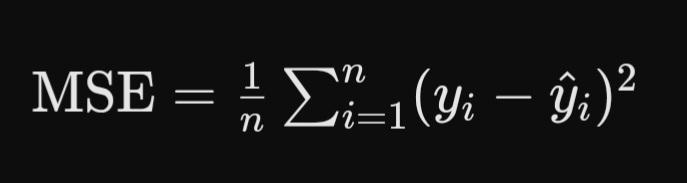


Linear Regression

Es tracta de trobar una relació lineal entre la variable d’entrada (x) i la variable de sortida (y). Serveix com a predictor de valors.

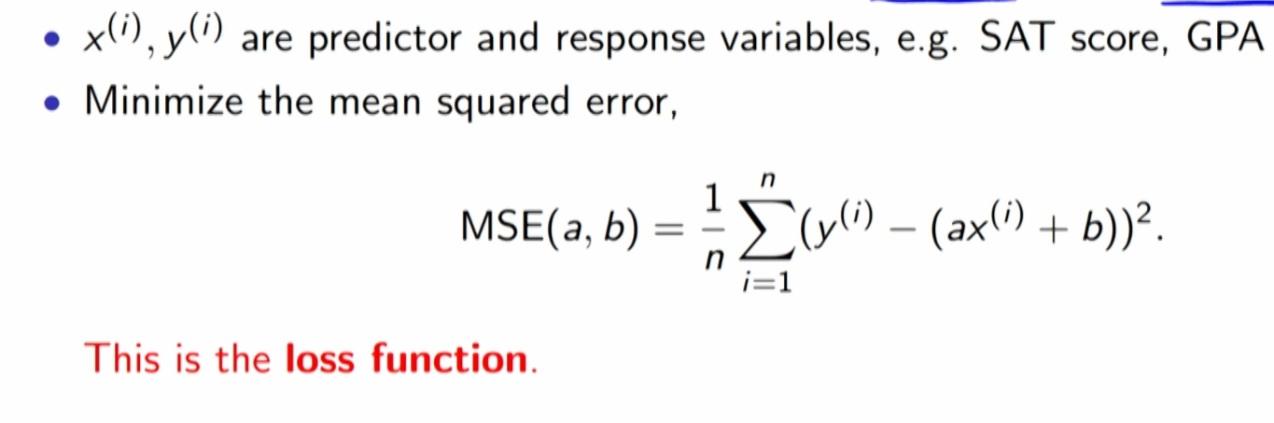


S’utilitza la mesura de l’error mitjà quadràtic per obtenir l’error mitjà en la predicció.



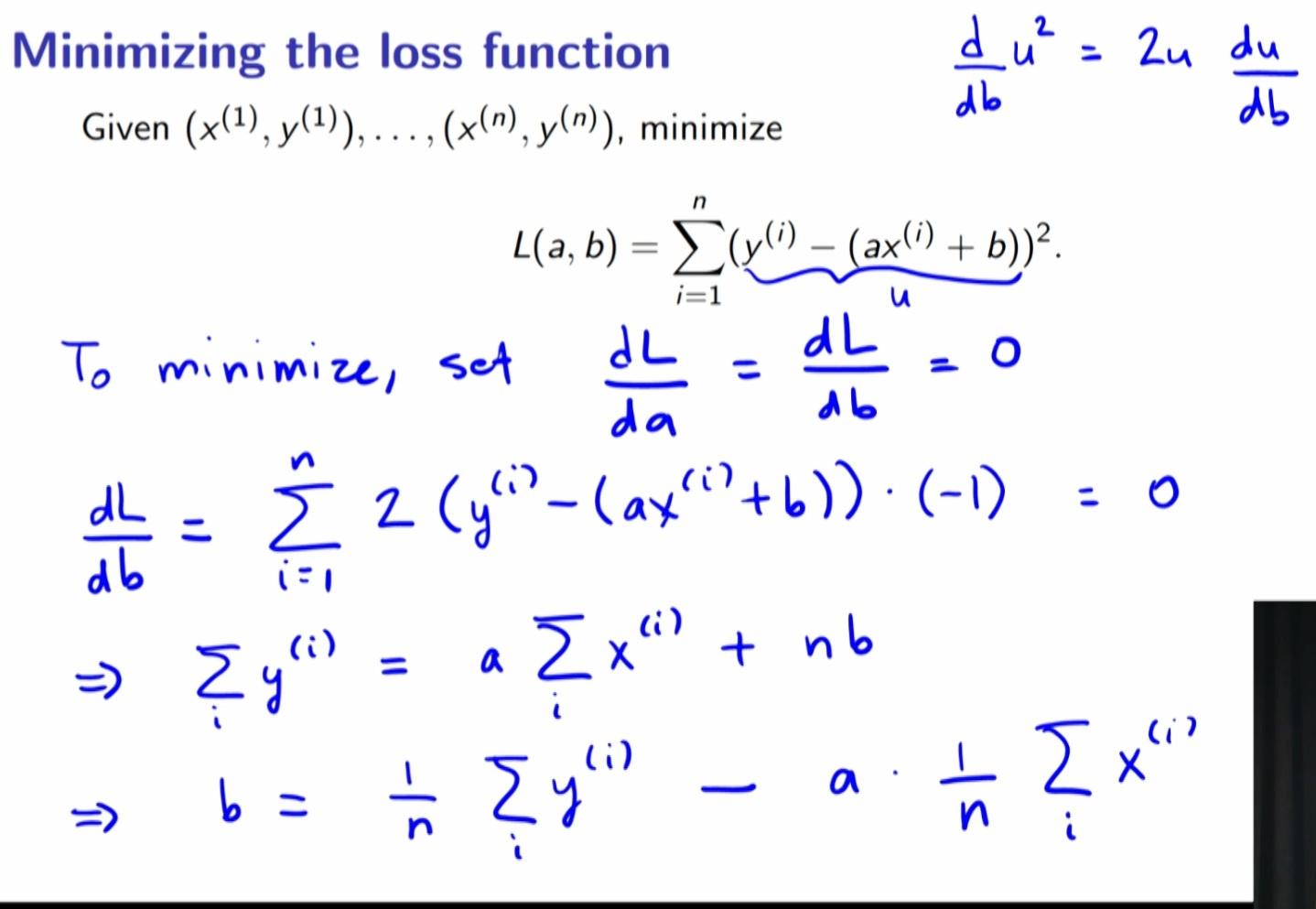
On y estimat és la mitjana obtinguda de les dades.

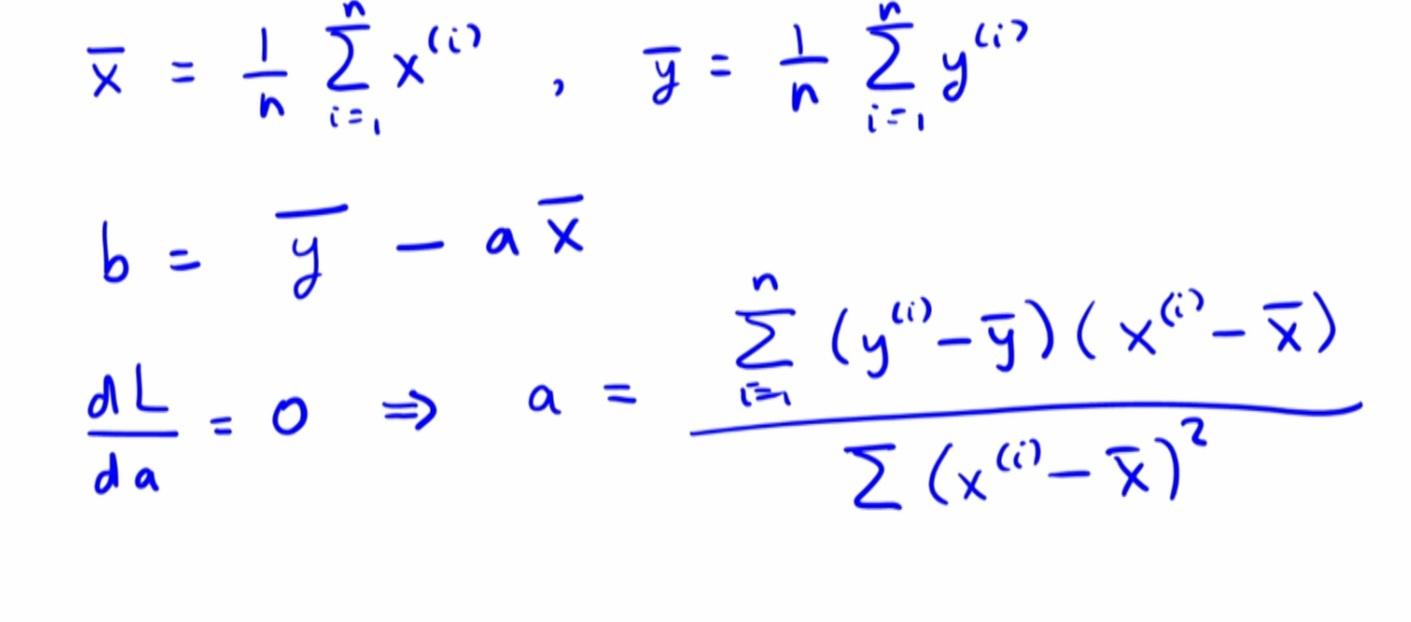
La recta té la forma y=ax+b, que s’inicialitza de forma arbitrària. Per optimitzar tant el pendent com el punt d’intercepció, s’utilitza una minimització de l’MSE:



a i b segueixen sent paràmetres desconeguts.

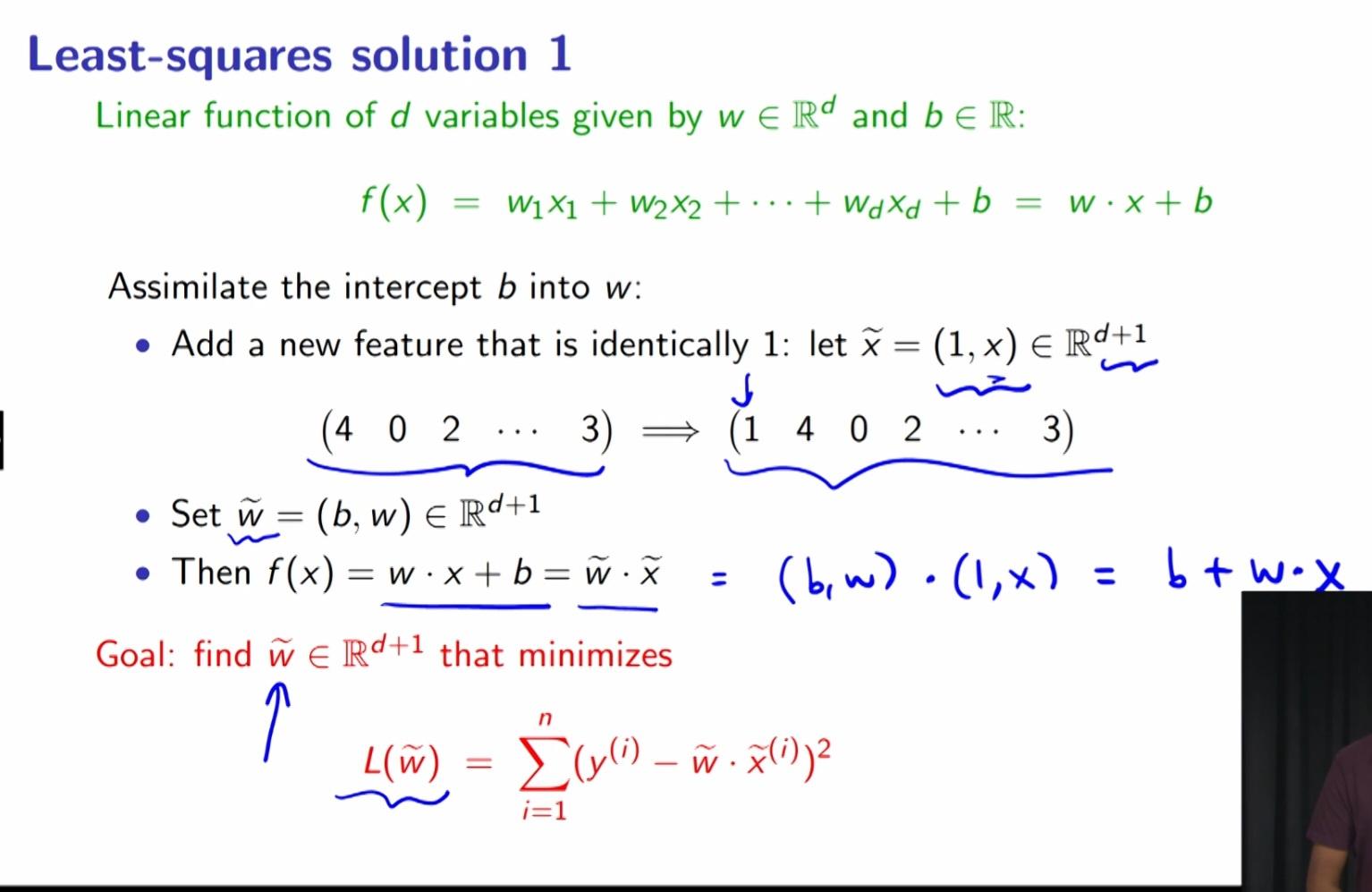
Mitjançant la derivada i obviant el terme 1/n, s'obté la minimització:

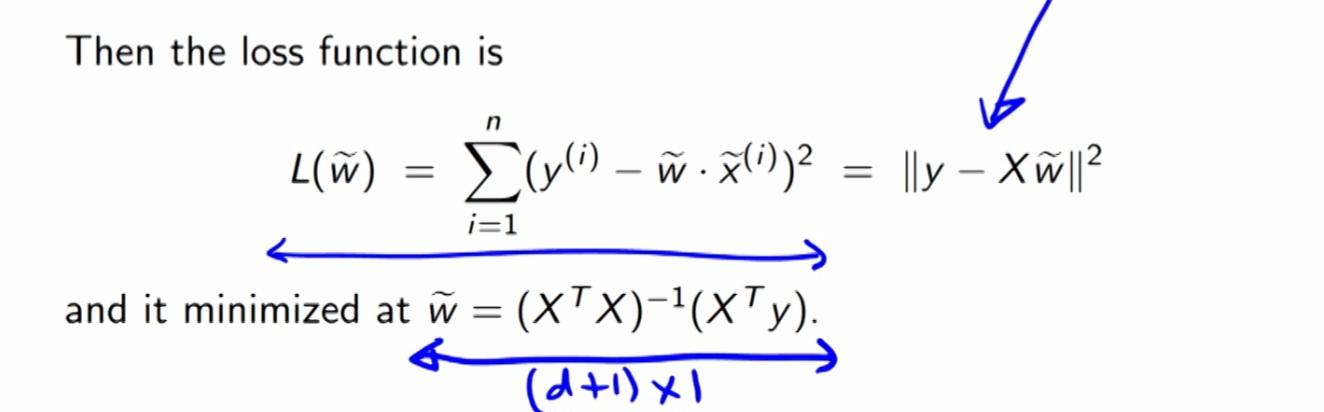




A la pràctica, s’utilitzen mètodes numèrics.

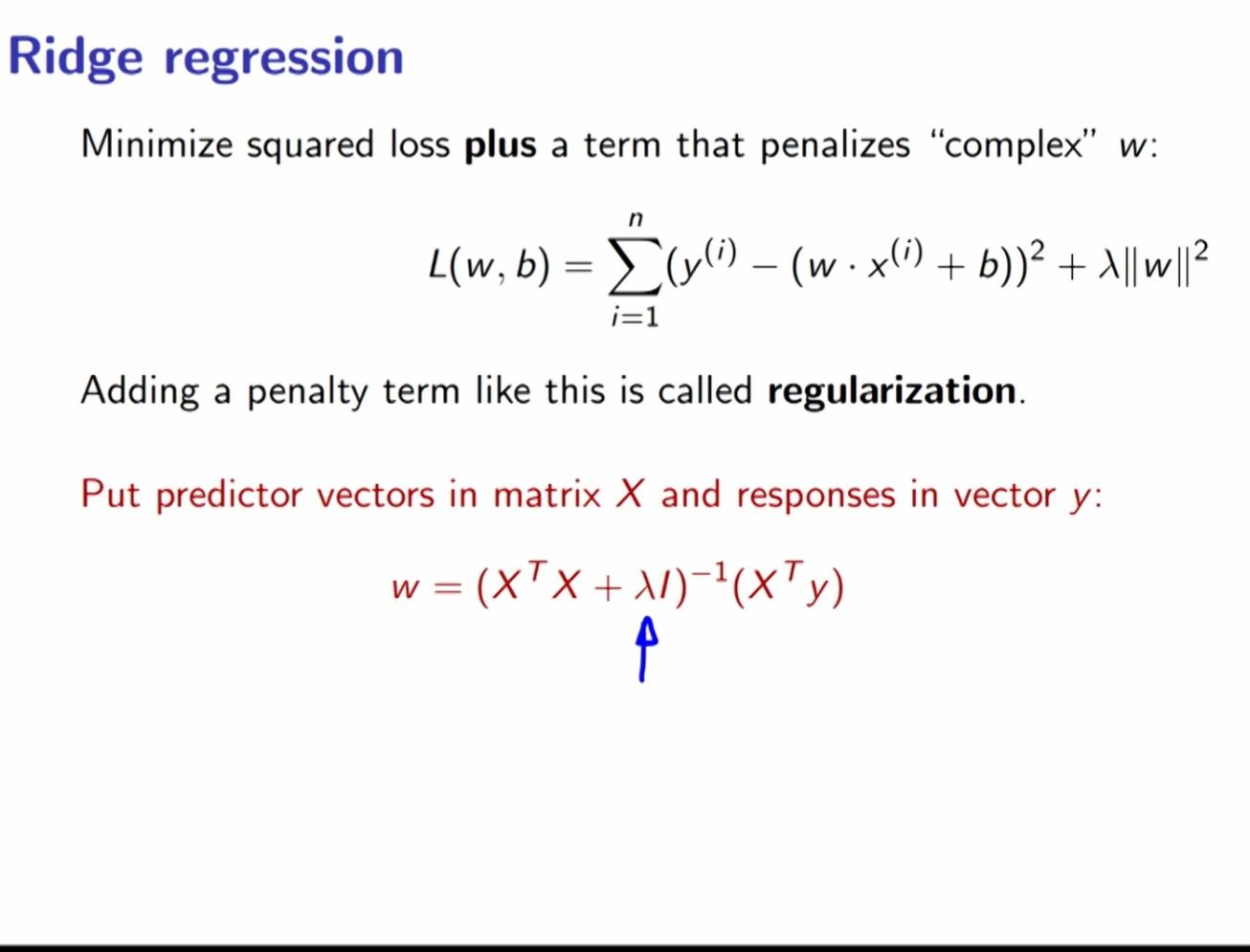
# Cas n-Dimensional





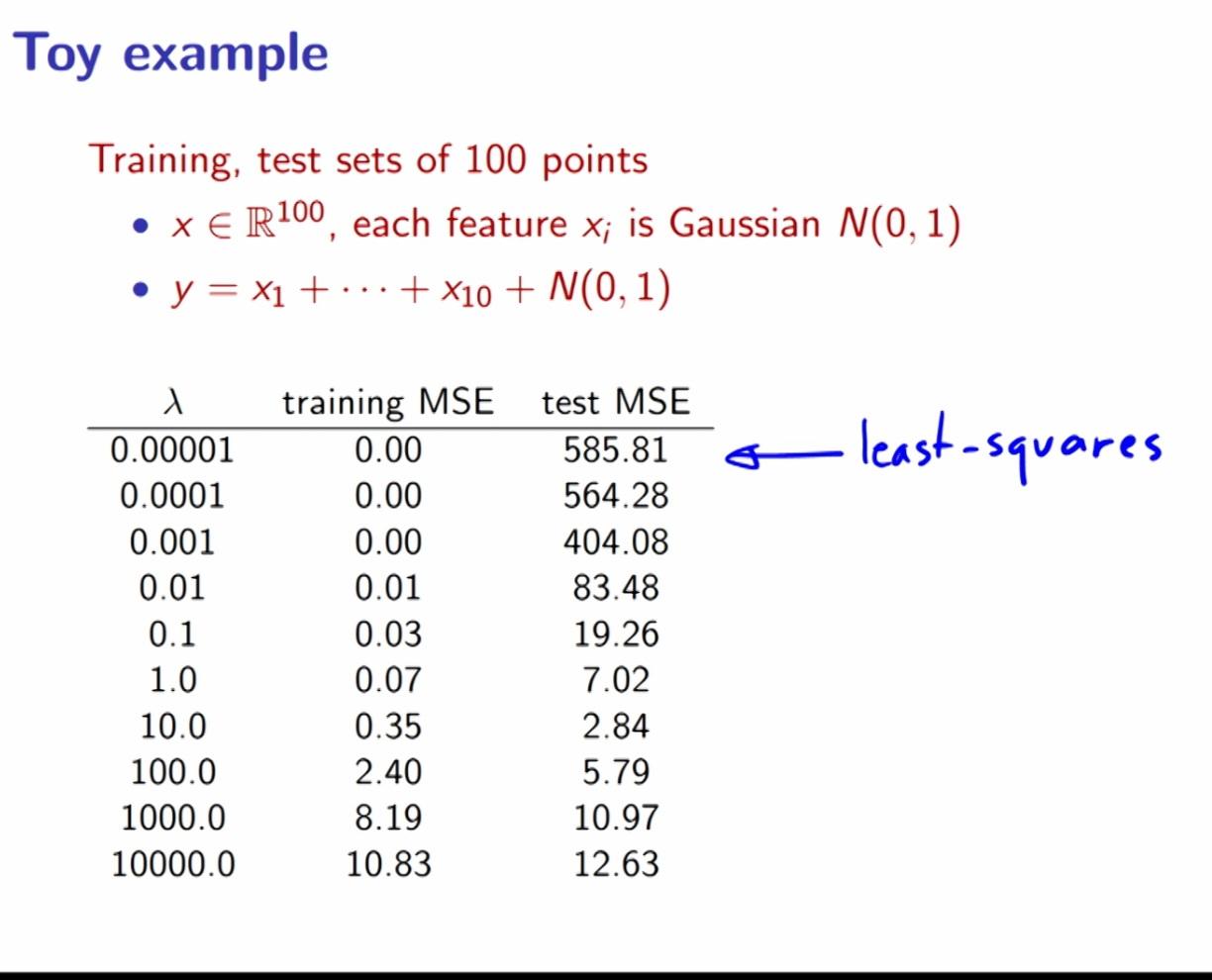
Ridge Regression

Afegeix a la Regressió Lineal un terme de penalització, lambda, que redueix la complexitat del vector de coeficients w, tal que:



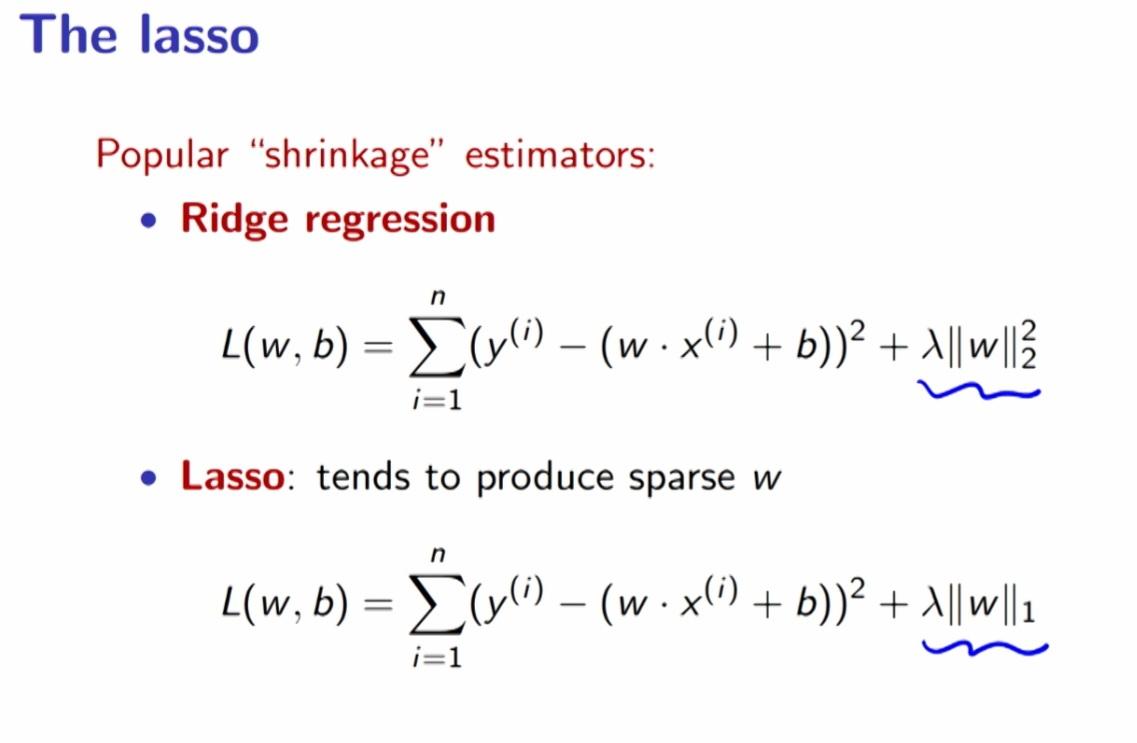
La matriu X és una matriu diagonal amb els valors predictors (x\_i) obtinguts de les dades experimentals, i al vector y els resultats experimentals.

De forma general, el terme de penalització ofereix millors resultats que els obtinguts amb el MSE:



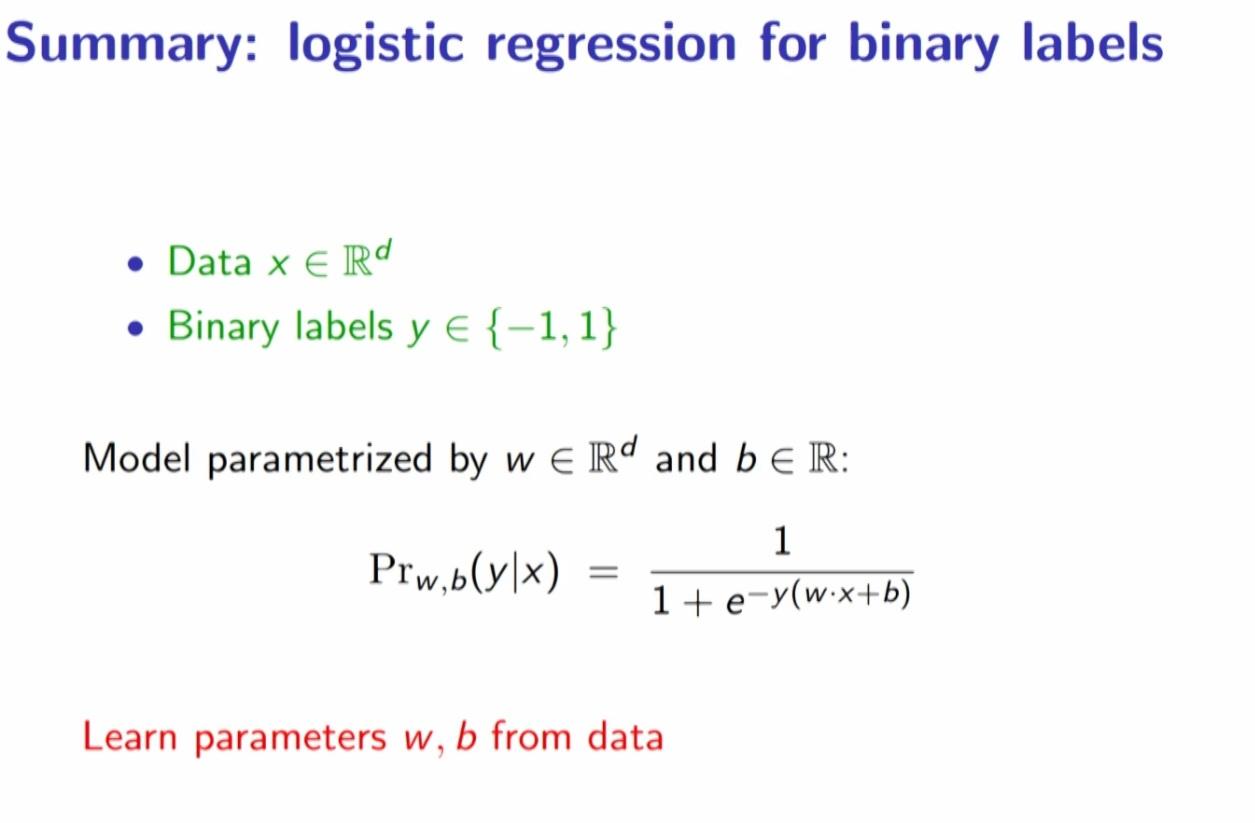
Lasso Regression

Un altre penalització molt utilitzada és utilitzar la norma l\_1 a l’hora d’afegir el terme de penalització. Aquesta formulació és la regressió lasso:



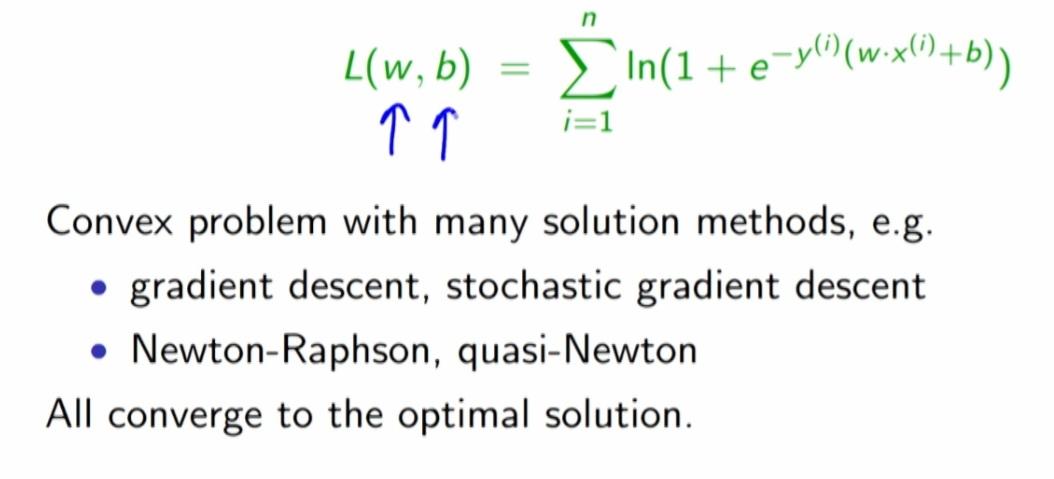
El subíndex 1 identifica la norma l\_1. Aquesta és utilitzada ja que dóna com a resultat la suma de les components al quadrat, que és major a la norma l\_2.

Logistic Regression



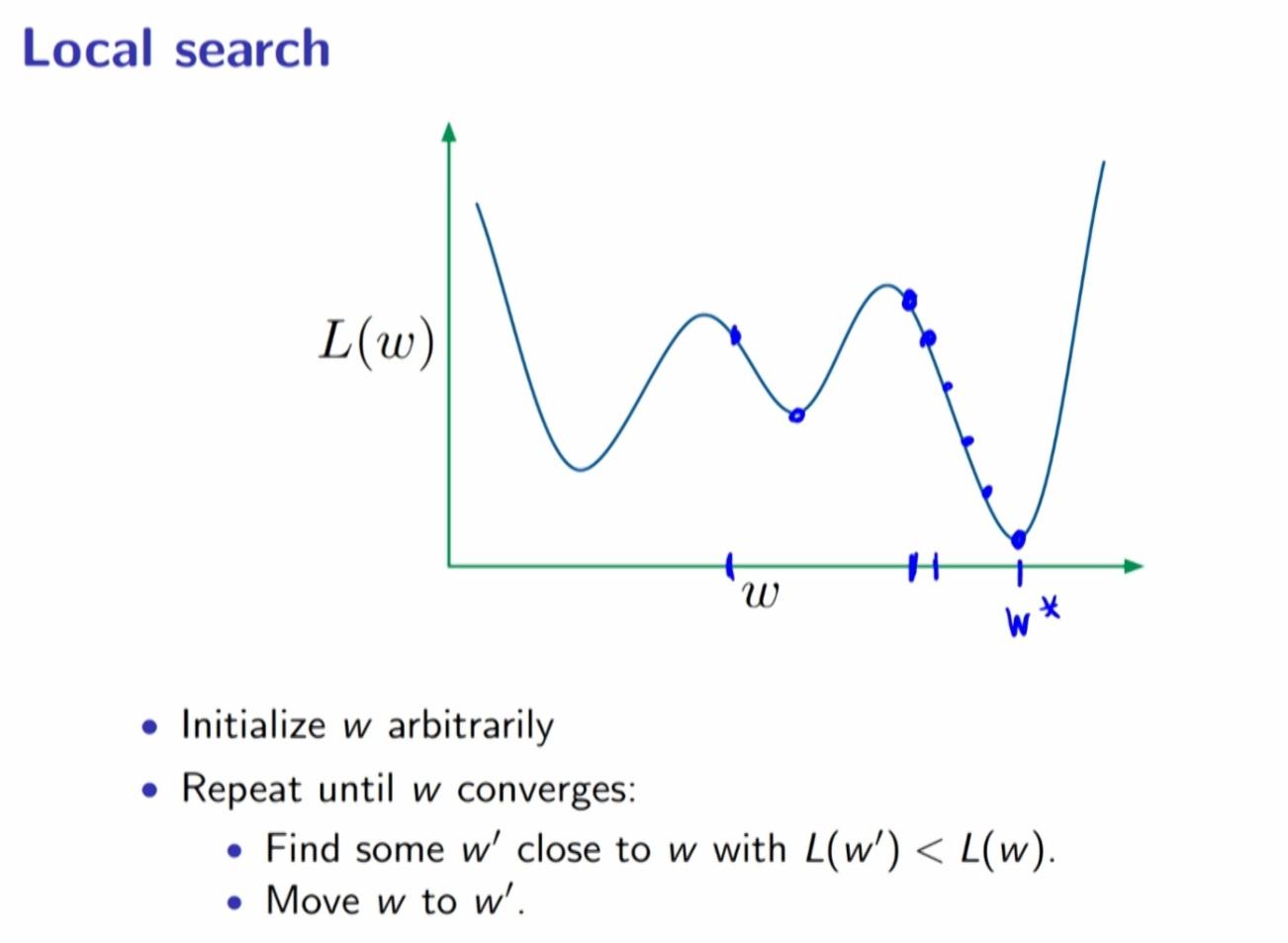
És d’utilitat per conèixer la probabilitat de la classificació. Primer s’obté **w** i **b** mitjançant una regressió lineal o penalitzada, i a partir d’aquestes dades es pot calcular la probabilitat de la classificació, coneixent així la confiança en que aquesta s’hagi dut a terme correctament.

Un altre mètode és tractar de minimitzar la funció logística per obtenir **w** i **b.**

****

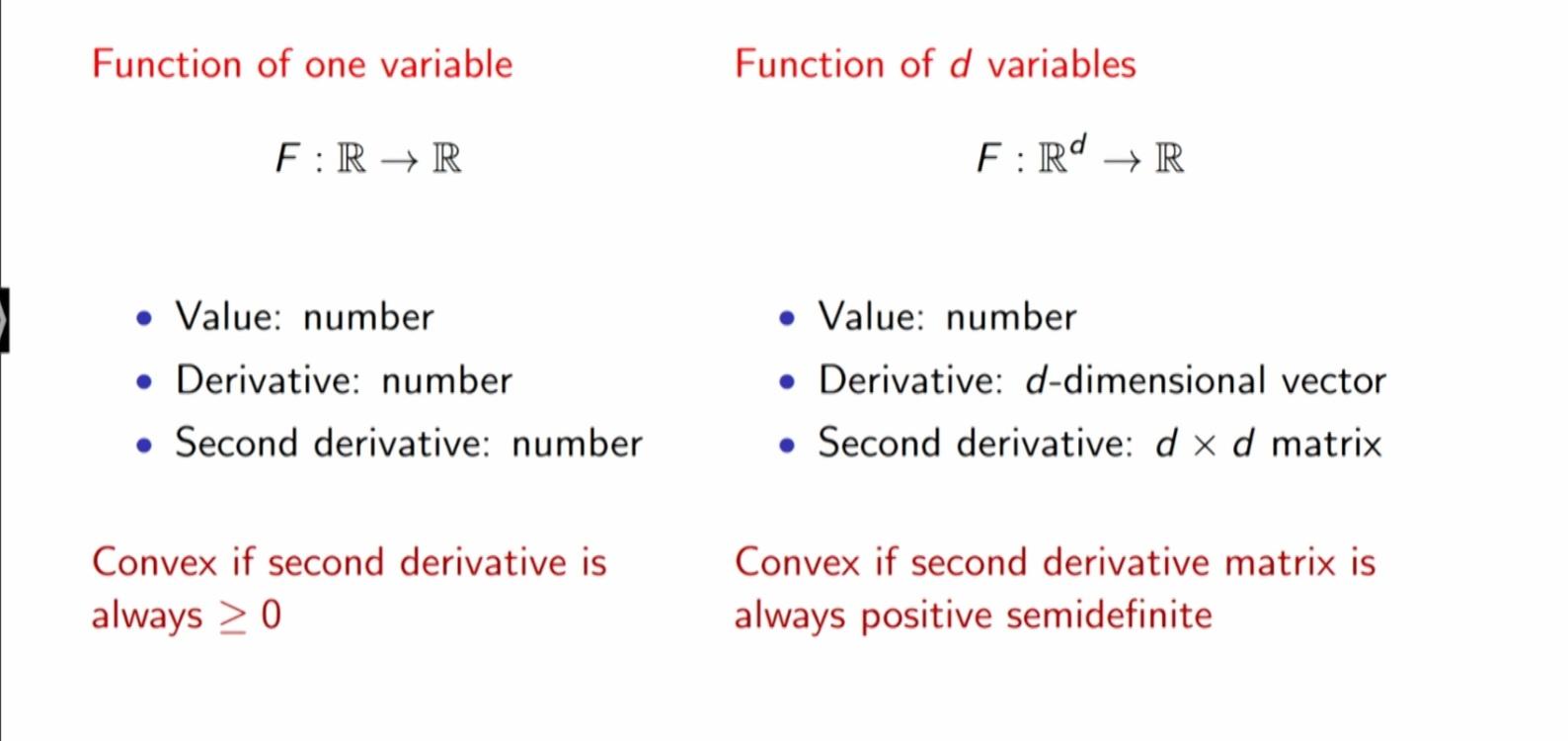
Algoritmes d’optimització

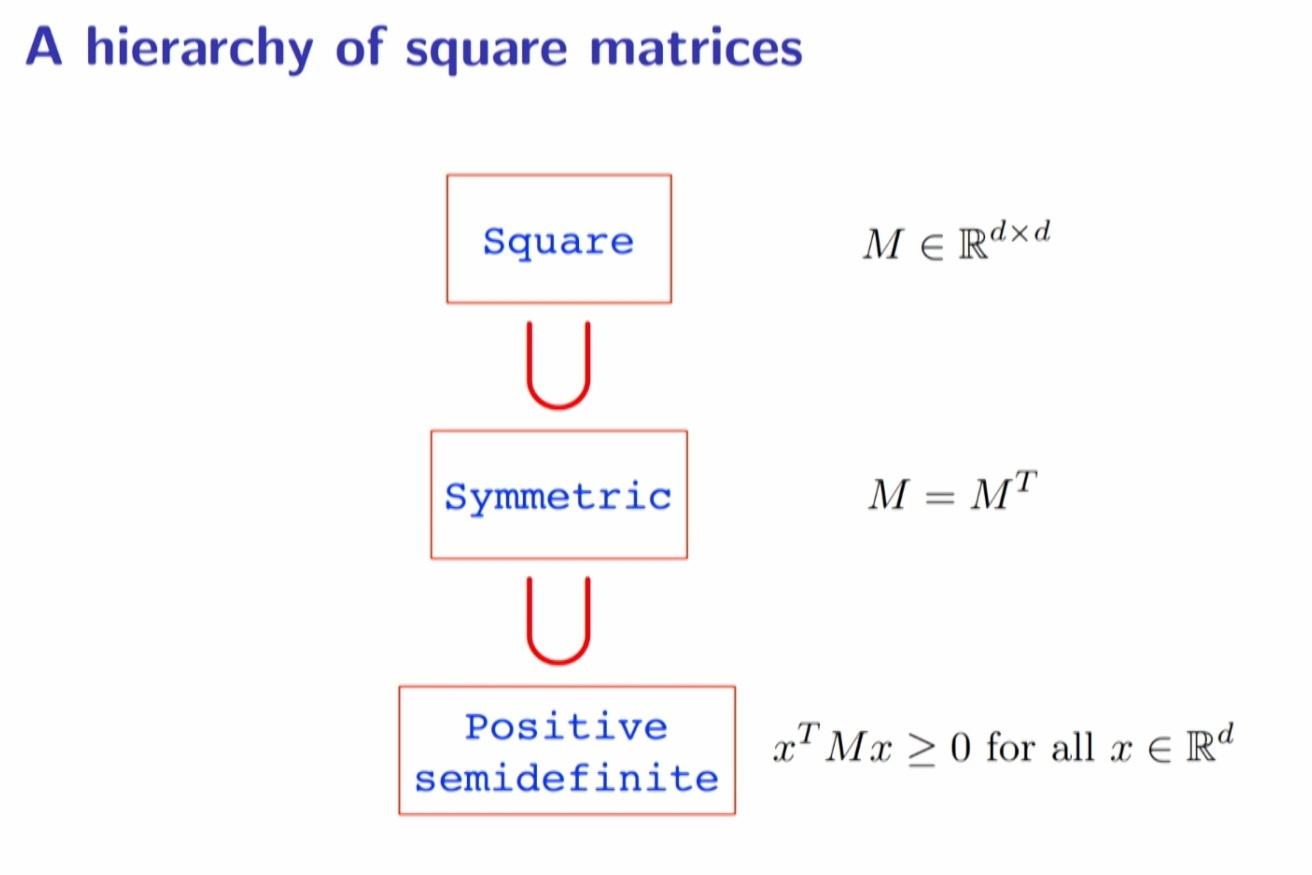
Local search



Búsqueda d’un mínim local en una funció. Funciona com a mínim absolut si la funció és convexa.

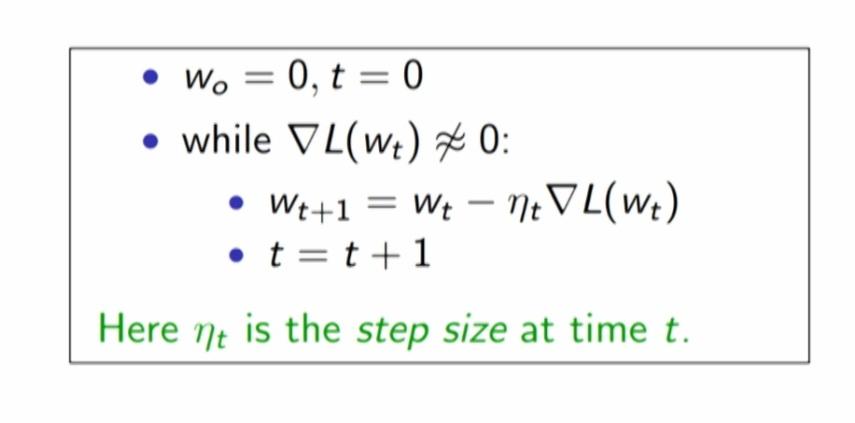
Per demostrar si la funció és convexa i poder aplicar el criteri de búsqueda local podent assegurar que la solució és òptima, s’aplica:



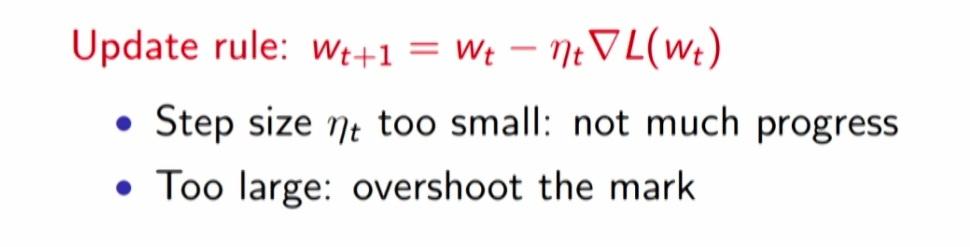


Gradient descent

És l’algoritme de cerca local per excel•lència, i utilitza com el seu nom indica el gradient per desplaçar el punt **w** en direcció on el gradient és mínim, assolint així el mínim local.

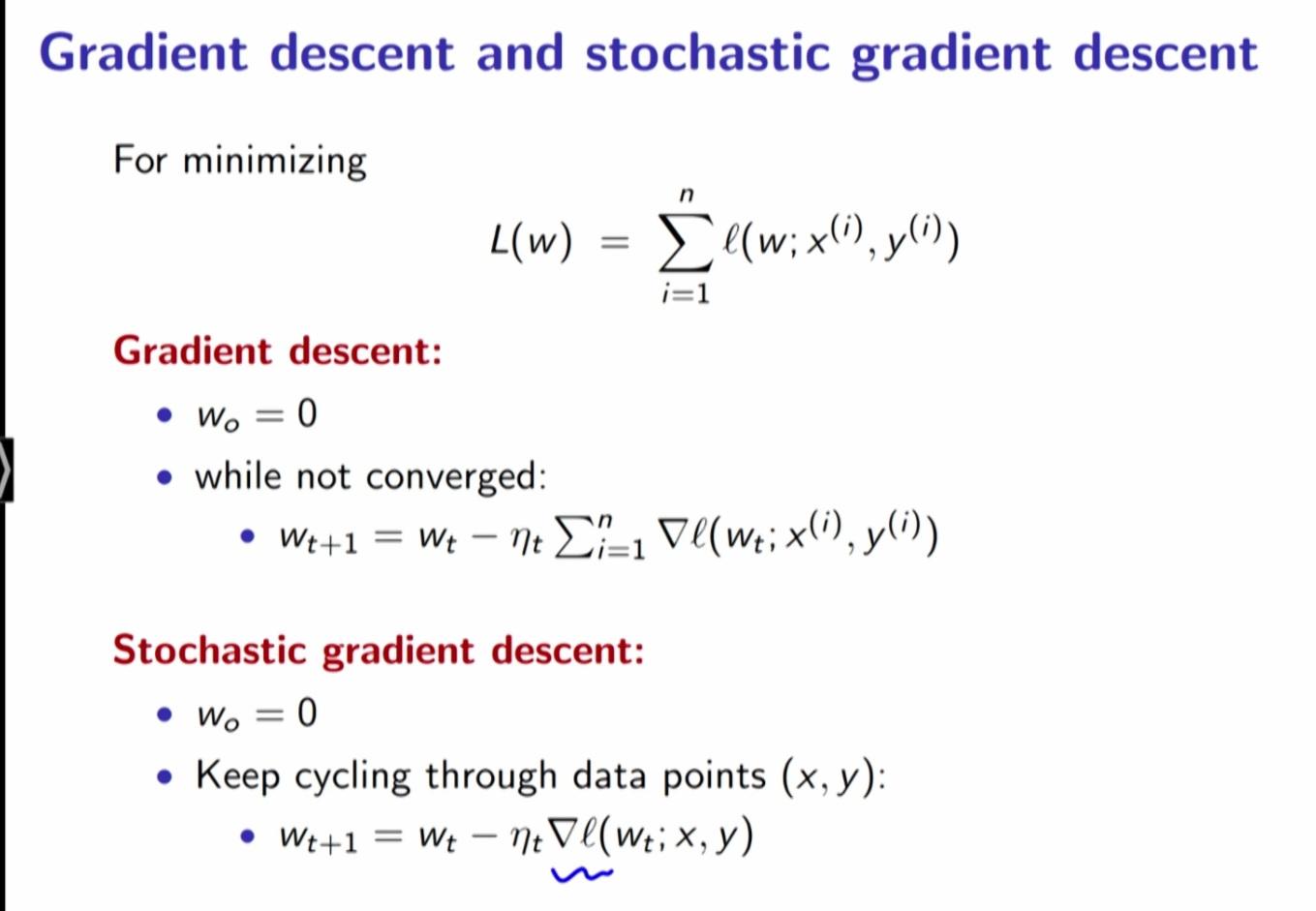


Eta\_t és un anàleg a la “resolució de la derivada”, ja que si aquest valor és molt petit, el salt de posició de w\_i a w\_i+1 és proporcional, fent que la cerca sigui més lenta però més precisa.



Stochastic descent

És una variació de Gradient descent que permet una búsqueda més eficient pero menys precisa.



En comptes de calcular la funció L(w) per tots els punts mostrals, per cada iteració del descens s’utilitza un punt de les mostres de forma cíclica. D’aquesta manera, en cada iteració de w es redueixen n-1 operacions.