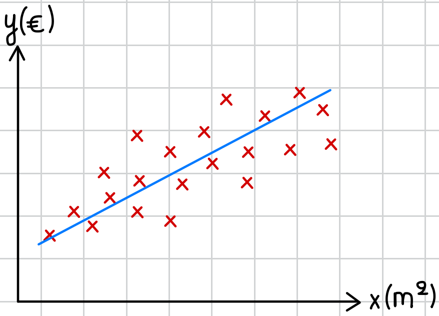
LINEAR REGRESSION

Πρόβλημα: Ένα μεσιτικό γραφείο έχει ζεύγη τετραγωνικά μέτρα - τιμή πώλησης για διάφορα σπίτια (τα οποία αναπαριστανται ως σημεία x στο επίπεδο) και επιθυμεί να βρει σχέση που θα υπολογίζει αυτόματα την τιμή πώλησης βάσει των τετραγωνικών μέτρων ενός σπιτιού.



Σκέψη: Θεωρούμε ότι η σχέση που συνδέει τα δύο μεγέθη είναι γραμμική, δηλαδή είναι μία σχέση της μορφής:

$\hat{y} = wx + b$

→ λανθασμένος όρος θόρυβος, bias

παράμετροι

→ βάρος, weight

(απλή γραμμική παλινδρόμηση, univariate linear regression)

Κριτήριο αξιολόγησης: Θέλουμε ένα μέτρο για τη διαφορά των προβλεπόμενων και των πραγματικών τιμών.

$$\underbrace{\hat{y}}_{\text{προβλεπόμενη}} - \underbrace{y}_{\text{πραγματική}} \longrightarrow (\hat{y}-y)^2 \longrightarrow \sum_{i=1}^{m}(\hat{y}^{(i)}-y^{(i)})^2 \longrightarrow \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)}-y^{(i)})^2 \longrightarrow \frac{1}{2}\cdot\frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)}-y^{(i)})^2$$

μεγαλύτερο σφάλμα => πολύ μεγαλύτερο penalty

m ζεύγη

ούτως ώστε να μη μεγαλώνει αυτόματα με το m

πιο neat

Αυτό ονομάζεται μέσο τετραγωνικό σφάλμα, RME και χρησιμοποιείται ως συνάρτηση κόστους, cost function

Σκοπός: Η επιλογή των παραμέτρων w, b ούτως ώστε να ελαχιστοποιηθεί η συνάρτηση κόστους J

Υπάρχει κλειστός μαθηματικός τύπος για τον σκοπό αυτό: normal equation (αυτό χρησιμοποιεί και η LinearRegression της Scikit-Learn)

GRADIENT DESCENT

Σκέψη: Κινούμαστε αντίθετα από την κλίση της J και οδηγούμαστε στο ελάχιστο της συνάρτησης. Τα w και b ανανεώνονται ως εξής:

$w = w - a\frac{\partial J}{\partial w}$ και $b = b - a\frac{\partial J}{\partial b}$

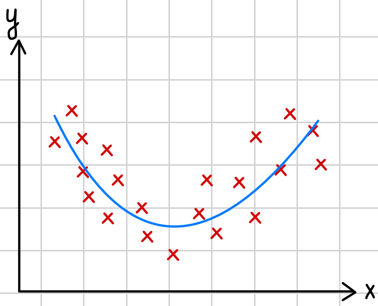
ρυθμός εκπαίδευσης, learning rate (μεγάλος, μικρός; τι γίνεται όταν είμαι στο ακρότατο;)

Γενικά βρίσκει τοπικό ελάχιστο. Για linear regression βρίσκει ολικό ελάχιστο.

Εάν είχαμε περισσότερα χαρακτηριστικά για το σπίτι: $\hat{y} = w_1x_1 + w_2x_2 + \ldots + w_nx_n + b$ (multivariate linear regression, πολυμεταβλητή γραμμική παλινδρόμηση)

POLYNOMIAL REGRESSION

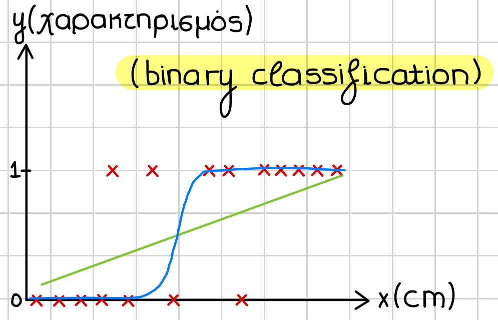
Πρόβλημα: Τα δεδομένα είναι πιο περίπλοκα από μια ευθεία γραμμή.



Σκέψη: Θεωρούμε ότι η σχέση που συνδέει τα δύο μεγέθη είναι πολυωνυμική, δηλαδή: $y = wx^n + \ldots + b$

LOGISTIC REGRESSION

Πρόβλημα: Ένα διαγνωστικό κέντρο έχει ζεύγη διάμετρος όγκου - χαρακτηρισμός (καλοήθης - κακοήθης) για διάφορους ανθρώπους (τα οποία αναπαρίστανται ως σημεία x στο επίπεδο) και επιθυμεί να βρει σχέση που θα υπολογίζει αυτόματα την σοβαρότητα βάσει της διαμέτρου.



Σκέψη: Η γραμμική σχέση δεν είναι καλή επιλογή για τα δεδομένα (βλ. πράσινο ευθύγραμμο τμήμα). Θέλουμε μία συνάρτηση με μορφή όπως η μπλε του σχήματος. Μία τέτοια συνάρτηση είναι η λογιστική ή σιγμοειδής·

$$\hat{p} = \frac{1}{1+e^{-z}} \qquad z = wx+b$$

(παλινδρόμηση με λογιστική συνάρτηση. logistic regression)

γιατί; Θέλουμε να πάμε από features στο [0,1]

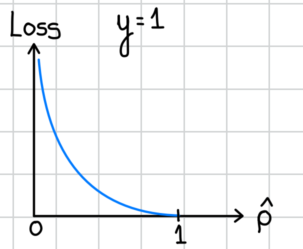
Το $\hat{p}$ έχει την έννοια της πιθανότητας. Μπορούμε τώρα να κάνουμε προβλέψεις ως εξής:

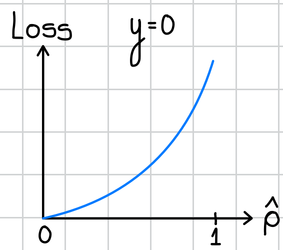
$$\hat{y} = \begin{cases} 0, & \hat{p} \leq p_0 \\ 1, & \hat{p} > p_0 \end{cases} \qquad p_0 \text{: threshold}$$

Κριτήριο αξιολόγησης: Η συνάρτηση κόστους που χρησιμοποιήσαμε στη γραμμική παλινδρόμηση πλέον έχει πολλά τοπικά ακρότατα, άρα χρησιμοποιήσουμε τη logistic loss.

$$\text{Loss} = \begin{cases} -\log(\hat{p}) & y=1 \\ -\log(1-\hat{p}) & y=0 \end{cases}$$

Ουσιαστικά δείχνει πόσο κοντά ή όχι είμαστε στην πραγματική ετικέτα. Όσο πιο μακρυά τόσο μεγαλύτερο penalty.





Άρα η συνάρτηση κόστους είναι η:

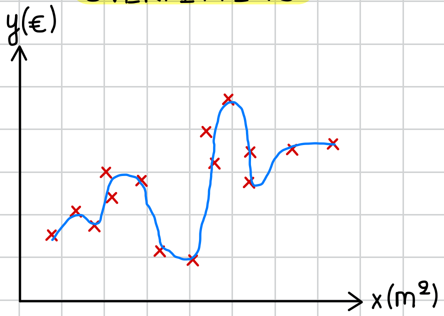
$$J = \frac{1}{m}\sum_{i=1}^{m} \text{Loss}(\hat{y}^{(i)}, y^{(i)})$$

Σκοπός: Η επιλογή των παραμέτρων w, b ούτως ώστε να ελαχιστοποιηθεί η συνάρτηση κόστους J.
Όχι κλειστός μαθηματικός τύπος, αλλά ίδια λογική με πριν.

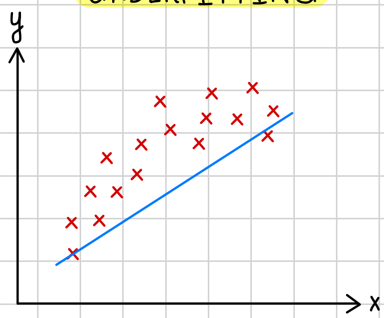
Εάν είχαμε παραπάνω από δύο κατηγορίες χρησιμοποιούμε (αντί για πολλούς binary classifiers) softmax regression (multinomial logistic regression)
Υπολογισμός πρώτα ενός score $s_k(x)$ για κάθε κατηγορία k και μετά υπολογισμός πιθανότητας με χρήση συνάρτησης softmax $\frac{e^{s_k(x)}}{\sum_{i=1}^{K} e^{s_i(x)}}$

Overfitting & Underfitting

OVERFITTING



UNDERFITTING



Είναι δυνατόν σπίτια με περισσότερα τετραγωνικά να έχουν μικρότερες τιμές πώλησης;

Θα ταίριαζε περισσότερο στα δεδομένα καμπύλη της μορφής x^2

Μπορούμε να κάνουμε καλύτερη επιλογή μοντέλου και διαχείριση δεδομένων.

Regularization

Συγκεκριμένα στην περίπτωση overfitting μπορούμε να χρησιμοποιήσουμε μία τεχνική που ονομάζεται regularization.

Ουσιαστικά περιορίζει τα βάρη w των features προσθέτοντας ένα επιπλέον penalty στη συνάρτηση κόστους.