ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

# ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

## 1$^{η}$ ΣΕΙΡΑ ΓΡΑΠΤΩΝ ΑΣΚΗΣΕΩΝ

Ειρήνη Δόντη
ΑΜ 03119839

7ο εξάμηνο

Αθήνα 2022

ΑΣΚΗΣΗ 1.1:

Για τη λύση της άσκησης χρησιμοποιήσαμε τη γλώσσα προγραμματισμού Python. Αρχικά, κανονικοποιούμε τα μεγέθη A, N και M με το γνωστό τύπο κανονικοποίησης $\frac{x-\mu}{\sigma}$ ώστε τα στοιχεία της εκάστοτε στήλης να έχουν μ=0 και σ=1.

(α) Υπολογίζουμε τον κανονικοποιημένο συντελεστή συσχέτισης r12 με χρήση της εντολής correlate(). Σχεδιάζουμε το γράφημα διασποράς για τις δοσμένες μεταβλητές και παρατηρούμε ότι τα σημεία των δοσμένων μεταβλητών είναι περισσότερα στην αρχή των αξόνων και τείνουν να ταυτιστούν με την ευθεία y=x.

(b) Linear Regression:

Υπολογίζουμε τα βάρη w0, w1 και w2 χωρίζοντας το σύνολο σε training set και test set. Οπότε, προκύπτει ότι w0 ≈ 0,00018, w1 ≈ 1,91266 και w2 ≈ -1,09768.

(c) Ridge Regression:

Υπολογίζουμε τα βάρη w0, w1 και w2 για τιμές παραμέτρου λ = {1, 10, 100}. Οπότε προκύπτει ότι για:

λ=1: w0 ≈ 0,00016, w1 ≈ 1,87293 και w2 ≈ -1,05814.

λ=10: w0 ≈2,83401, w1 ≈ 1,59128 και w2 ≈ -0,77830.

λ=100: w0 ≈ -0,00027, w1 ≈ 0,80214 και w2 ≈ -0,01098.

(d) Υπολογίζουμε τα σφάλματα RMSE εκπαίδευσης και επαλήθευσης που αντιστοιχούν στα παραπάνω ερωτήματα (b) και (c) με χρήση της εντολής sqrt(mean_squared_error()). Συμπληρώνουμε τον παρακάτω πίνακα:

| | Linear Regression λ=0 | Ridge Regression λ=1 | Ridge Regression λ=10 | Ridge Regression λ=100 |
|---|---|---|---|---|
| RMSE(Train Set) | 0,49650 | 0,49658 | 0,50155 | 0,55344 |
| RMSE (Test Set) | 0,47601 | 0,47565 | 0,47691 | 0,51423 |
| Εξίσωση y= | $0{,}00018+1{,}91266x_1-1{,}09768x_1$ | $0{,}00016+1{,}87293x_1-1{,}05814x_2$ | $2{,}83401+1{,}59128x_1-0{,}77830x_2$ | $-0{,}00027+0{,}80214x_1-0{,}01098x_2$ |

Η τιμή της παραμέτρου λ που θα επιλέγαμε είναι εκείνη της οποίας ο μέσος όρος των RMSE μεταξύ των εκάστοτε train και test set) είναι ο μικρότερος. Οπότε, βάσει των παραπάνω, θα επιλέγαμε την παράμετρο λ=0.

Παρατηρούμε ότι, καθώς αυξάνεται η παράμετρος λ, τότε οι απόλυτες τιμές των w0, w1 και w2 βαρών πλησιάζουν περισσότερο την τιμή 0.

ΑΣΚΗΣΗ 1.2:

(a) Ισχύει ότι: E[x] = $\int_{-\infty}^{\infty} x\, p(x)dx = \int_{-\infty}^{\infty} x\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx =$

$= \int_{-\infty}^{\infty} (y+\mu)\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y)^2}{2\sigma^2}}dy = \int_{-\infty}^{\infty} y\ \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y)^2}{2\sigma^2}}dy + \mu \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y)^2}{2\sigma^2}}dy$ = μ

Επίσης, $\sigma_x^2$ = E[x$^2$] – E[x] = $\sigma^2 + \mu^2$ - $\mu^2$ = $\sigma^2$ διότι

E[x$^2$] = $\int_{-\infty}^{\infty} x^2\, p(x)dx = \int_{-\infty}^{\infty} x^2\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx =$

$= \int_{-\infty}^{\infty} (y+\mu)^2\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y)^2}{2\sigma^2}}dy = \int_{-\infty}^{\infty} y^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y)^2}{2\sigma^2}}dy +$

$+ \mu^2 \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y)^2}{2\sigma^2}}dy + 2\mu \int_{-\infty}^{\infty} y \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y)^2}{2\sigma^2}}dy = \sigma^2 + \mu^2$

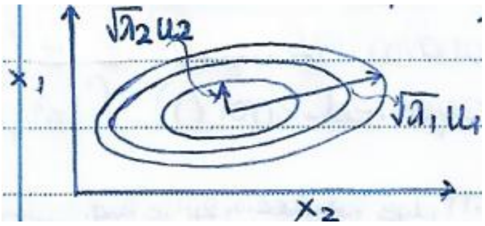
(b) Γενίκευση κανονικής κατανομής:

p(x) = $\frac{1}{(2\pi)^{\frac{l}{2}}|\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}$

Για να βρούμε τις ισοσταθμικές καμπύλες πρέπει p(x) = c ή

$(x-\mu)^T\Sigma^{-1}(x-\mu) = c$ ή $y^T\Lambda^{-1}y$ = [y1 y2] $\begin{bmatrix} \frac{1}{\lambda 1} & 0 \\ 0 & \frac{1}{\lambda 2} \end{bmatrix} \begin{bmatrix} y1 \\ y2 \end{bmatrix} = \frac{y1^2}{\lambda 1^2} + \frac{y2^2}{\lambda 2^2}$ = c.

Οι ισοσταθμικές είναι ελλείψεις.




(c) Εκτελούμε τη μέθοδο μέγιστης πιθανοφάνειας

L(μ) = ln($\prod_{n=1}^{N} P(xn-\mu)$) = ln($\frac{1}{(\sigma\sqrt{2\pi})^N} e^{-\frac{\Sigma(x-\mu)^2}{2\sigma^2}}$) =

= -N(ln($\sigma\sqrt{2\pi}$)) - $\sum_{n=1}^{N} \frac{(xn-\mu)^2}{2\sigma^2}$

Πρέπει: $\frac{\partial L(\mu)}{\partial \mu}$ = 0 ή $\sum_{n=1}^{N}(xn-\mu)^2 = 0$ ή $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} xn$.

ΑΣΚΗΣΗ 1.3:

(a) Πρέπει: P(ω1) P(xt|ω1) = P(ω2) P(xt|ω2) ή $\frac{1}{2}\frac{1}{2a}\exp(-\frac{|xt|}{\alpha}) = \frac{1}{2}\frac{1}{2a}\exp(-\frac{|xt-3|}{\alpha})$ ή xt = $\frac{3}{2}$

(b) Πρέπει: λ12P(ω1) P(xr|ω1) = λ21P(ω2) P(xr|ω2) ή $\frac{1}{2}\frac{1}{2}\frac{1}{2a}\exp(-\frac{|xr|}{\alpha}) = \frac{1}{2}\frac{1}{2a}\exp(-\frac{|xr-3|}{\alpha})$ ή xr = $\frac{3-ln2}{2}$

Ισχύει ότι xr < xt, το οποίο είναι λογικό, καθώς προστίθενται στη (b) περίπτωση τα βάρη λ12, λ21.

(c) (i) Η εξίσωση της καμπύλης απόφασης:

$(\mu1^T\Sigma^{-1} - \mu2^T\Sigma^{-1})x' - \frac{1}{2}\mu1^T\Sigma^{-1}\mu1 + \frac{1}{2}\mu2^T\Sigma^{-1}\mu2 = 0$ ή

([2 -2]$\begin{bmatrix}\frac{2}{3} & 0\\ 0 & \frac{2}{3}\end{bmatrix}$- [-1 2]$\begin{bmatrix}\frac{2}{3} & 0\\ 0 & \frac{2}{3}\end{bmatrix}$)x' - $\frac{1}{2}$ [2 -2]$\begin{bmatrix}\frac{2}{3} & 0\\ 0 & \frac{2}{3}\end{bmatrix}\begin{bmatrix}2\\ -2\end{bmatrix}$ + $\frac{1}{2}$ [-1 2]$\begin{bmatrix}\frac{2}{3} & 0\\ 0 & \frac{2}{3}\end{bmatrix}\begin{bmatrix}-1\\ 2\end{bmatrix}$=0

ή 6x1 – 8x2 – 3 =0.

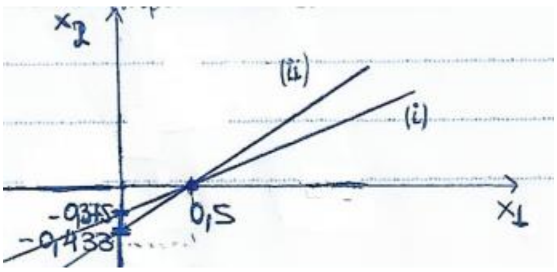
(ii) Η εξίσωση της καμπύλης απόφασης:

$(\mu1^T\Sigma^{-1} - \mu2^T\Sigma^{-1})x' - \frac{1}{2}\mu1^T\Sigma^{-1}\mu1 + \frac{1}{2}\mu2^T\Sigma^{-1}\mu2 = 0$ ή

([2 -2]$\begin{bmatrix}\frac{3}{4} & -\frac{1}{4}\\ -\frac{1}{4} & \frac{3}{4}\end{bmatrix}$ - [-1 2]$\begin{bmatrix}\frac{3}{4} & -\frac{1}{4}\\ -\frac{1}{4} & \frac{3}{4}\end{bmatrix}$)x' - $\frac{1}{2}$ [2 -2]$\begin{bmatrix}\frac{3}{4} & -\frac{1}{4}\\ -\frac{1}{4} & \frac{3}{4}\end{bmatrix}\begin{bmatrix}2\\ -2\end{bmatrix}$ +

+ $\frac{1}{2}$ [-1 2]$\begin{bmatrix}\frac{3}{4} & -\frac{1}{4}\\ -\frac{1}{4} & \frac{3}{4}\end{bmatrix}\begin{bmatrix}-1\\ 2\end{bmatrix}$ = 0 ή 26x1 – 30x2 – 13 =0

Παρακάτω παραθέτουμε τη γεωμετρική απεικόνιση των παραπάνω ευθειών:



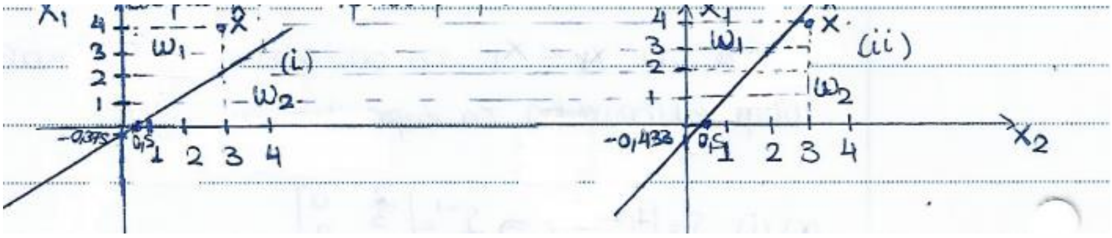

Παρατηρούμε ότι οι και δύο εξισώσεις καμπύλης απόφασης διέρχονται απ' το ($\frac{1}{2}$, 0).

(d) Αλγεβρικός προσδιορισμός:

(i) 6x1 – 8x2 – 3 = -3 < 0. Οπότε το πρότυπο θα ταξινομηθεί στην κλάση ω1.

(ii) $26x1 - 30x2 - 13 = 1 > 0$. Οπότε το πρότυπο θα ταξινομηθεί στην κλάση ω2.

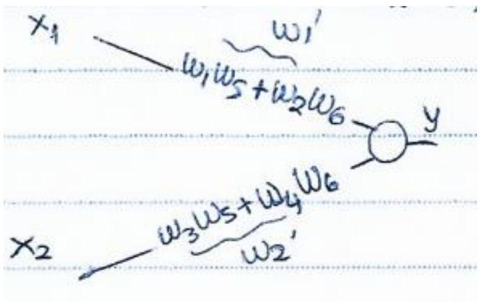
Γεωμετρικός Προσδιορισμός:




Από τα παραπάνω διαγράμματα παρατηρούμε ότι, για την ευθεία διαχωρισμού (i), το πρότυπο x’ βρίσκεται πάνω από εκείνη και συνεπώς θα ταξινομηθεί στην κλάση ω1. Επίσης, για την ευθεία διαχωρισμού (ii) το πρότυπο x’ βρίσκεται κάτω από εκείνη και συνεπώς θα ταξινομηθεί στην κλάση ω2.

ΑΣΚΗΣΗ 1.4:

(a) $y=f[f(x1w1 + x2w3)]w5 + [f(x1w2+x2w4)]w6] = f[c(x1w1+x2w3)w5 + c(x1w2+x2w4)w6] = c^2((w1w5+w2w6)x1 + (w3w5+w4w6)x2) = c^2u'$




(b)Από το παραπάνω ερώτημα, γενικεύεται ότι η έξοδος του MLP με κ-κρυφά επίπεδα μπορεί να γραφτεί ως $c^{k+1}(w1x1+w2x2+\ldots+wnxn)$ όπου w1, w2, … ,wn οι συναρτήσεις αρχικών βαρών. Οπότε, μπορούμε να βρούμε το ισοδύναμο νευρωνικό δίκτυο χωρίς κρυφά επίπεδα με γραμμική συνάρτηση ενεργοποίησης $f(u) = c^{k+1}u$.

(c) $y=g[f(x1w1 + x2w3)]w5 + [f(x1w2+x2w4)]w6] =$
$g[\frac{w5}{1+e^{-(x1w1+x2w3)}} + \frac{w6}{1+e^{-(x1w2+x2w4)}}]$.

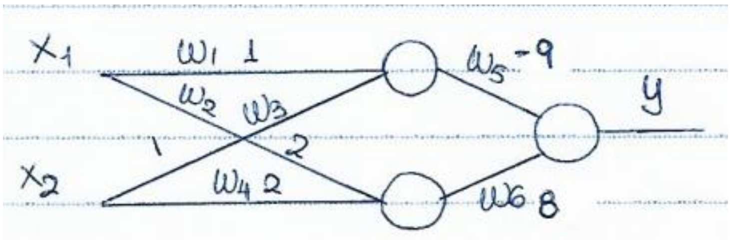
Για x1=0, x2=1: $u'= \frac{w5}{1+e^{-(w3)}} + \frac{w6}{1+e^{-(w4)}} > 0$

Για x1=1, x2=0: u’= $\frac{w5}{1+e^{-(w1)}} + \frac{w6}{1+e^{-(w2)}} > 0$

Για x1=0, x2=0: u’= $\frac{w5}{2} + \frac{w6}{2} < 0$

Για x1=1, x2=1: u’= $\frac{w5}{1+e^{-(w1+w3)}} + \frac{w6}{1+e^{-(w2+w4)}} < 0$

Από τα παραπάνω: 0,731w5+0,881w6 > 0 και w5+w6 < 0. Οπότε, w1=1, w3=1, w2=2, w4=2, w5=-9 και w6=8.

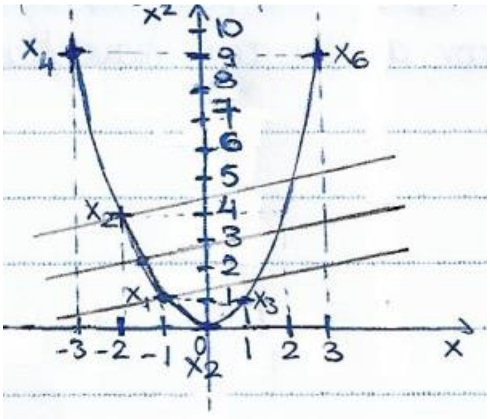



<u>ΑΣΚΗΣΗ 1.5:</u>

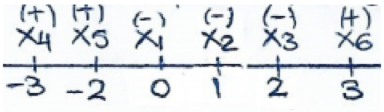
(a) Μορφή πίνακα K(X1,X1’) = X1X1’(1+X1X1’)

(b) Παρατηρούμε ότι η γραμμή πρέπει να περάσει περίπου ανάμεσα στα σημεία u1=(-2,4) και u2=(-1,1). Το παραπάνω ισχύει, καθώς αποτελούν τα πλησιέστερα σημεία της αντίθετης τάξης. Το περιθώριο γίνεται μέγιστο στην περίπτωση που το όριο απόφασης πρέπει να περνά από το κέντρο ($-\frac{3}{2}, \frac{5}{2}$) της γραμμής που συνδέονται τα σημεία (-2,4) και (-1,1). Επίσης, πρέπει να είναι κάθετο στο ευθύγραμμο τμήμα που συνδέει τα σημεία (-2,4) και (-1,1). Οπότε, το όριο απόφασης μπορεί να περιγραφεί, κάνοντας χρήση της εξίσωσης: -y1+3y2-9=0. Το πλάτος γ του περιθωρίου είναι η απόσταση μεταξύ ενός απ’τα κανονικά επίπεδα και του διαχωριστικού υπερεπιπέδου, δηλαδή $\frac{\sqrt{10}}{2}$.

(c)




(d) Όριο απόφασης του υπερεπιπέδου διαχωρισμού (R’)




(e) Λαμβάνουμε υπόψη το δυϊκό πρόβλημα τετραγωνικού προγραμματισμού στα στα διανύσματα u1 και u2. Οπότε, μόνο τα a1 και a2 είναι μη μηδενικά. Επίσης, γνωρίζουμε ότι a1=a2=a από τους περιορισμούς του τετραγωνικού προγράμματος.

Οπότε: $y'(x) = \sum_{n=1}^{N} an + \frac{1}{2}\sum_{n=1} \sum_{i=1}^{N} an\ ai\ tn\ ti\ K(xn, xi) = 2a - 5a^2$.

Πρέπει $y'(x) = 0$ ή α = α1 = α2 = $\frac{1}{5}$.

Σε θετικό περιθώριο, ισχύει ότι: 1 = $\sum_{n=1}^{|SV|} an\ yn\ K(x, un) + b$ ή b = -$\frac{9}{5}$.

(f) Όχι, καθώς το σημείο που προστέθηκε, βρίσκεται εκτός περιθωρίου και συνεπώς ταξινομείται σωστά.

(g) Ισχύει ότι K(u,v) = $\sum_{i=1}^{\infty} \frac{e^{-\frac{u^2}{2}} u^i}{\sqrt{i!}} \frac{e^{-\frac{v^2}{2}} v^i}{\sqrt{v!}} = \exp\left(-\frac{u^2+v^2}{2}\right) \sum_{i=1}^{N} \frac{(uv)^i}{i!} =$

$\exp\left(-\frac{u^2+v^2}{2}\right) \exp(uv) = \exp\left(-\frac{(u-v)^2}{2}\right)$.

(h) Γενικά, είναι άξιο ανησυχίας, αλλά στην περίπτωση των SVM η πολυπλοκότητα μοντέλου δεν αντιμετωπίζεται απ’τη διάσταση του χώρου των χαρακτηριστικών. Αντίθετα, αντιμετωπίζεται απ’τον αριθμό των διανυσμάτων υποστήριξης.

ΑΣΚΗΣΗ 1.6:

(a) 1. Ισχύει ότι: $\sum_{i=1}^{v} a_i \cdot log\, xi \ \le log(\sum_{i=1}^{v} a_i\, xi)$ με $a_i \ge 0$, $\sum_{i=1}^{v} a_i$ =1,

P(k) = $\sum_{x} P(x,k)$, P(x) = $\sum_{k} P(x,k)$.

Οπότε, ig(x,k) = H(x) – H(x|k) =

= - $\sum_{x} P(x)\log_2(P(x)) - \sum_{k} P(k) \sum_{x} (-P(x|k)\log_2(P(x|k)))$ ή

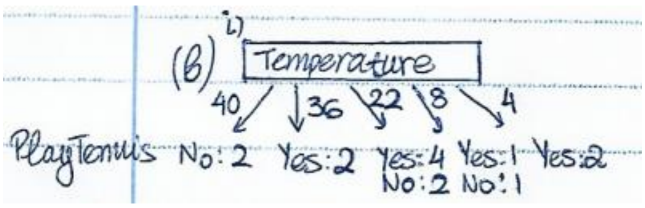
ig(x,k) = - $\sum_{x} \sum_{k} P(x,k)((\log_2(P(x)) - \log_2(P(x|k))) \ge$

$\sum_{k} P(x)(\log_2(\sum_{x} \frac{P(x|k)P(x)}{P(x|k)}))$ ή ig(x,k) ≥ -( $\log_2\left(\sum_{k} P(k) \sum_{x} P(x)\right)$

ή ig(t)≥0.

2. Εφόσον εφαρμόζεται έλεγχος κριτηρίου ισότητας που αφορά την ίδια επιλογή χαρακτηριστικού, τότε τα ti και tj δεν είναι ανεξάρτητα. Οπότε, ισχύει ότι H[ti] – H[ti|tj] ≠ 0 και συνεπώς ισχύει ότι ig(tj) > 0 βάσει ερωτήματος 1.

(b) (i)




gini(Temperature) = $\frac{2}{14}$gini(Temperature = 40) + $\frac{2}{14}$gini(Temperature=36) + $\frac{6}{14}$gini(Temperature=22) + $\frac{2}{14}$gini(Temperature=8) + $\frac{2}{14}$gini(Temperature=4) = $\frac{11}{42}$.

gini(Temperature=40) = 1-$(\frac{0}{2})^2 - \left(\frac{2}{2}\right)^2 = 0$,

gini(Temperature=36) = 1-$(\frac{2}{2})^2 - \left(\frac{0}{2}\right)^2 = 0$

gini(Temperature=22) = 1-$(\frac{4}{6})^2 - \left(\frac{2}{6}\right)^2 = \frac{4}{9}$

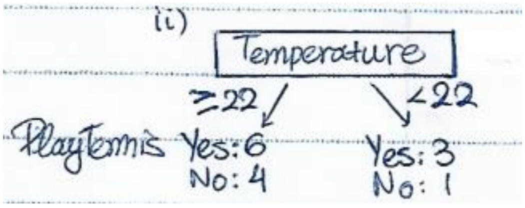
gini(Temperature=8) = 1-$(\frac{1}{2})^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2}$

gini(Temperature=4) = 1-$(\frac{2}{2})^2 - \left(\frac{0}{2}\right)^2 = 0$

gini(Root) = 1-$(\frac{9}{14})^2 - \left(\frac{5}{14}\right)^2 = \frac{45}{98}$

Οπότε, ig(Temperature) = gini(Root) – gini(Temperature) = $\frac{45}{98}$ - $\frac{11}{42}$.= 0,1973

(ii)




gini(Temperature) = $\frac{10}{14}$gini(Temperature ≥ 22) + $\frac{4}{14}$gini(Temperature < 22) = $\frac{69}{196}$.

gini(Temperature ≥ 22) = 1-$(\frac{6}{10})^2 - \left(\frac{4}{10}\right)^2 = \frac{12}{35}$

gini(Temperature < 22) = 1-$(\frac{3}{4})^2 - \left(\frac{1}{4}\right)^2 = \frac{3}{8}$

gini(Root) = 1-$(\frac{9}{14})^2 - \left(\frac{5}{14}\right)^2 = \frac{45}{98}$

Οπότε, ig(Temperature) = gini(Root) – gini(Temperature) = $\frac{45}{98}$ - $\frac{69}{196}$.= 0,1071

Το δέντρο που θα επιλέγαμε είναι το πρώτο, καθώς το κέρδος είναι μεγαλύτερο σε σχέση με το δεύτερο.