Χατζηθεοδώρου Ιάσων 03117089

Αναγνώριση Προτύπων

1η Αναλυτική Εργασία

el17089@mail.ntua.gr

Άσκηση 1.1

Η κατανομή των δειγμάτων είναι η

$$p(x|\theta) = \theta^2 x e^{-\theta x} u(x)$$

Άρα περνώντας το λογάριθμο

$$lnp(x|\theta) = 2ln\theta - \theta x + lnx \quad για\ x \geq 0$$

Παραγωγίζοντας ως προς θ

$$\frac{\partial \ln p(x|\theta)}{\partial \theta} = \frac{2}{\theta} - x$$

Άρα προκύπτει η παρακάτω εξίσωση για τη μεγιστοποίηση της πιθανοφάνειας

$$\sum_{i=1}^{n} \frac{2}{\theta} - x_i = 0$$

Η οποία λύνεται ως προς θ ως εξής

$$\theta = \frac{2n}{\sum x_i}$$

Άσκηση 1.2

Η εξίσωση για τη μέθοδο minimax είναι η εξής

$$\lambda_{11} + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(x|\omega_1)dx = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(x|\omega_2)dx$$

Αντικαθιστώντας τις τιμές $\lambda_{11} = \lambda_{22} = 0$ και $\lambda_{12} = \lambda_{21} = 1$ έχουμε

$$\int_{R_2} p(x|\omega_1)dx = \int_{R_1} p(x|\omega_2)dx$$

Η παραπάνω λύση δεν είναι μοναδική ως προς τις δύο κατανομές. Σαν αντιπαράδειγμα μπορούμε να σκεφτούμε οποιεσδήποτε δύο κατανομές που περιορίζονται η κάθε μία στην περιοχή της. Άρα και τα δύο ολοκληρώματα προκύπτουν μηδέν και ισχύει η σχέση όμως οι κατανομές μπορούν να είναι οποιεσδήποτε.

Άσκηση 1.3

Στις δύο διαστάσεις η πιθανότητα σφάλματος είναι

$$P[2d\ error] = \int_{D_{11}} \int_{D_{12}} p(x_1, x_2|\omega_2) p(\omega_2) dx_2 dx_1 + \int_{D_{21}} \int_{D_{22}} p(x_1, x_2|\omega_1) p(\omega_1) dx_2 dx_1$$

όπου προφανώς τα D_{ij} ορίζουν τις περιοχές απόφασης κάθε κατηγορίας. Για να υπολογίσουμε την προβολή ολοκληρώνουμε ως προς x_2, οπότε η πιθανότητα σφάλματος στη μία διάσταση είναι

$$P[1d\ error] = \int_{D_1} \int_{-\infty}^{\infty} p(x_1, x_2|\omega_2) dx_2 * p(\omega_2) dx_1 + \int_{D_2} \int_{-\infty}^{\infty} p(x_1, x_2|\omega_1) dx_2 * p(\omega_1) dx_1$$

όπου αντίστοιχα τα D_i ορίζουν τις περιοχές απόφασης. Τα D_i όμως δεν μπορούν να μην περιέχουν τα D_{i1}, γιατί στις δύο διαστάσεις οι περιοχές έχουν επιλεγεί έτσι ώστε να ελαχιστοποιείται η πιθανότητα σφάλματος χρησιμοποιώντας περισσότερη πληροφορία από ότι στη μία διάσταση. Άρα προκύπτουν οι εξής σχέσεις

$$\int_{-\infty}^{\infty} p(x_1, x_2|\omega_2) dx_2 * p(\omega_2) = \int_{-\infty}^{\infty} p(x_1, x_2|\omega_2) p(\omega_2) dx_2 \geq \int_{D_{12}} p(x_1, x_2|\omega_2) p(\omega_2) dx_2$$

επειδή η πιθανότητα είναι πάντα μη αρνητική και

$$\int_{D_1} \int_{-\infty}^{\infty} p(x_1, x_2|\omega_2) dx_2 * p(\omega_2) dx_1 \geq \int_{D_{11}} \int_{D_{12}} p(x_1, x_2|\omega_2) p(\omega_2) dx_2 dx_1$$

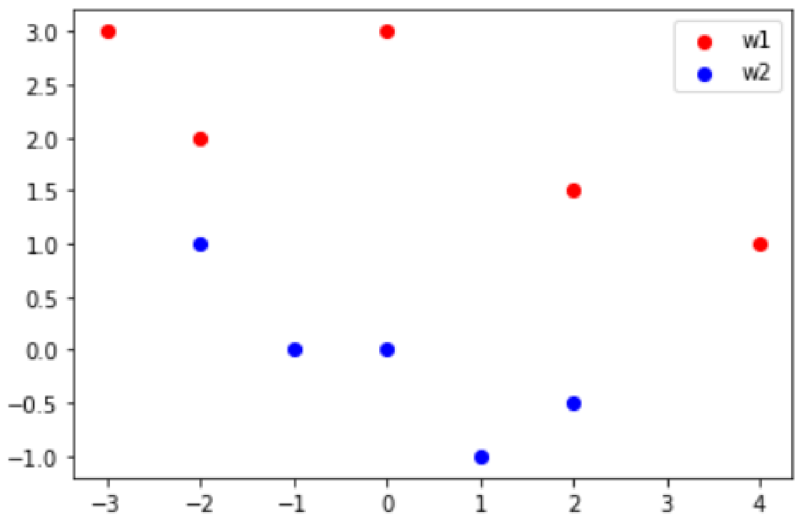
επειδή $D_{11} \subseteq D_1$. Ομοίως για το ολοκλήρωμα στη περιοχή του ω_2. Άρα τελικά ισχύει

$$P[1d\ error] \geq P[2d\ error]$$

Το παραπάνω αποτέλεσμα ισχύει και στην κανονική κατανομή ως ειδική περίπτωση.

Άσκηση 1.4

Σχεδιάζοντας τα σημεία προκύπτει η εξής εικόνα, από την οποία είναι προφανές πως οι κλάσεις είναι γραμμικά διαχωρίσιμες

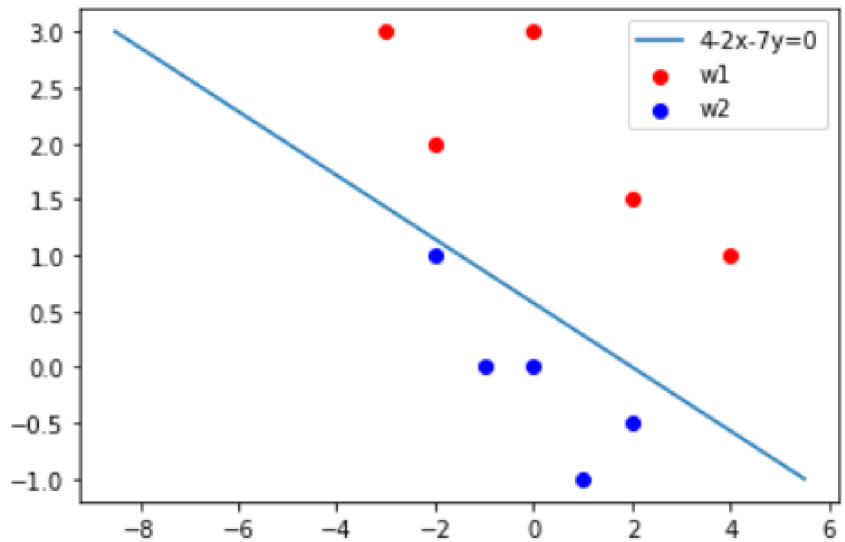


Για διάνυσμα βαρών $w = [0, 0]$ το perceptron από ένα σημείο και μετά κάνει πάντα λάθος στα δείγματα (-2, 1) και (2, -0.5). Αυτό συμβαίνει γιατί χωρίς το w_0, η ευθεία που προσπαθεί να βρει περνάει από την αρχή των αξόνων και όπως φαίνεται στο σχήμα μια τέτοια ευθεία δεν θα μπορούσε να διαχωρίσει τις δύο κλάσεις.

Για $w = [0, 0, 0]$ επαυξάνουμε τα δείγματα βάζοντας το 1 στην πρώτη θέση τους. Παρακάτω φαίνονται τα updates που γίνονται στις πρώτες εποχές

```
 i    f(w(t)*x(i))  Correction        w(t)              w(t+1)
---  --------------  ----------------  ----------------  ----------------
 0               1  [-1.  -2.  -1.5]  [0 0 0]           [-1.  -2.  -1.5]
 2               1  [-1.  3. -3.]     [-1.  -2.  -1.5]  [-1.  3. -3.]
 5               0  [ 1. -1.  0.]     [-1.  3. -3.]     [ 1. -1.  0.]
 7               0  [ 1.   2.  -0.5]  [ 1. -1.  0.]     [ 1.   2.  -0.5]
 9               0  [ 1. -2.  1.]     [ 1.   2.  -0.5]  [ 1. -2.  1.]


 i    f(w(t)*x(i))  Correction     w(t)           w(t+1)
---  --------------  -------------  -------------  -------------
 2               1  [-1.  3. -3.]  [ 1. -2.  1.]  [ 0.  1. -2.]
 5               0  [-1.  3. -3.]  [ 0.  1. -2.]  [ 0.  1. -2.]
 9               0  [-1.  3. -3.]  [ 0.  1. -2.]  [ 0.  1. -2.]


 i    f(w(t)*x(i))  Correction     w(t)           w(t+1)
---  --------------  -------------  -------------  -------------
 1               1  [-1. -4. -1.]  [ 0.  1. -2.]  [-1. -3. -3.]
 6               0  [1. 0. 0.]     [-1. -3. -3.]  [ 1.  1. -2.]
 9               0  [ 1. -2.  1.]  [ 1.  1. -2.]  [ 1. -1. -1.]
```

Μετά από 14 εποχές ολοκληρώνεται η εκτέλεση και προκύπτουν τα βάρη $w = [4, -2, -7]$ οπότε η ευθεία είναι η $4 - 2x - 7y = 0$ όπως φαίνεται και στη γραφική παράσταση



Παρατίθεται ο κώδικας του perceptron

```
def f(vect, weight):
  res = np.round(np.dot(vect, weight), decimals=2)
  if res < 0:
    return 0
  else:
    return 1

def one_cycle(weight, x, y, b):
  stuff = []
  temp_w = weight
  for i in range(len(x)):
    estimate = f(x[i], temp_w)
    if y[i] == estimate:
      continue
    if np.round(np.dot(weight, x[i]), 2) <= 0 and y[i] == 1:
      weighted_x = b * x[i]
    elif np.round(np.dot(weight, x[i]), 2) >= 0 and y[i] == 0:
      weighted_x =  - b * x[i]

    new_w = weight + weighted_x
    stuff.append([i, estimate, weighted_x, temp_w, new_w])
    temp_w = new_w
  return (stuff, temp_w)
```

```
def train(weight, x_train, y_train, b_param, max_epochs=5):
  stuff = []
  for i in range(max_epochs):
    (stuff, weight) = one_cycle(weight, x_train, y_train, b_param)
    if not stuff:
      break
    print(tabulate(stuff, headers=['i', 'f(w(t)*x(i))', 'Correction', 'w(t
)', 'w(t+1)']))
    print('\n')
  return weight
```

Άσκηση 1.5

Επιλέγω να αντικαταστήσω τη δεδομένη κατανομή στην p_2 και να θεωρήσω ότι άγνωστη είναι η p_1 οπότε

$$D_{KL} = \int p_1(x) ln p_1(x) dx - \int p_1(x) ln p_2(x) dx$$

$$= \int p_1(x) ln p_1(x) dx + \frac{1}{2}\int p_1(x)[d ln(2\pi) + \ln|\Sigma| + (x-\mu)^T \Sigma^{-1}(x-\mu)]dx$$

Για την ελαχιστοποίηση αρκεί να πάρω τις παραγώγους ως προς μ και Σ^{-1} ως εξής

$$\frac{\partial D_{KL}}{\partial \mu} = -\int \Sigma^{-1}(x-\mu) p_1(x) dx = 0$$

Η παραπάνω σχέση απλοποιείται σε

$$\int p_1(x)(x-\mu)dx = E[x-\mu] = 0 \Rightarrow \mu = E[x]$$

Παραγωγίζοντας ως προς Σ^{-1}

$$\frac{\partial D_{KL}}{\partial \Sigma^{-1}} = \int p_1(x)[-\Sigma + (x-\mu)(x-\mu)^T]dx = 0$$

Η σχέση απλοποιείται σε

$$E[\Sigma - (x-\mu)(x-\mu)^T] = 0 \Rightarrow \Sigma = E[(x-\mu)(x-\mu)^T]$$

Άσκηση 1.6

Ερώτημα 1

Το σφάλμα του linear regression προκύπτει από τη διαφορά μεταξύ της σωστής τιμής και της πρόβλεψης δηλαδή

$$error_i = y_i - w^T x_i$$

Αν υποθέσουμε ότι αυτό το σφάλμα ακολουθεί κανονική κατανομή $N(0, \sigma^2)$ τότε η λύση μέγιστης πιθανοφάνειας προκύπτει ως εξής

$$lnL = \sum \{\ln(2\pi\sigma^2) + \frac{-(yi - w^T x_i)^2}{2\sigma^2}\}$$

Η συγκεκριμένη ποσότητα μεγιστοποιείται όταν ελαχιστοποιείται το

$$\sum (yi - w^T x_i)^2$$

Όμως αυτό είναι ακριβώς η λύση που ψάχνει ο αλγόριθμος LMS. Άρα οι λύσεις είναι ισοδύναμες.

Ερώτημα 2

Ψάχνουμε ευθεία της μορφή $y = b + ax$ οπότε χρησιμοποιώντας τη μέθοδο ελάχιστων τετραγώνων έχουμε τους εξής τύπους

$$a = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2},$$

$$b = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Κάνοντας τις πράξεις προκύπτει η ευθεία $y = 1.228 + 2.204x$

$Ερώτημα\ 3$

Ο αλγόριθμος LMS κάνει update τα βάρη με βάση τον παρακάτω κανόνα

$$w(t+1) = w(t) + r * (y_n - w(t)^T x_n) * x_n$$

Για την συγκεκριμένη εκτέλεση αρχικοποιώ το $w(0) = [1, 1]$ και επιλέγω $r = 0.1$ και η εκτέλεση φαίνεται στον πίνακα

x_n	y_n	w	$r(y_n - w(t)^T x_n)x_n$
0.38	2.05	[1, 1]	[0.067, 0.025]
0.44	2.23	[1.067, 1.025]	[0.071, 0.031]
0.48	2.13	[1.14, 1.056]	[0.048, 0.023]
0.54	2.33	[1.186, 1.08]	[0.056, 0.03]
0.58	2.67	[1.242, 1.11]	[0.078, 0.045]
0.64	2.68	[1.32, 1.155]	[0.062, 0.039]
0.71	2.81	[1.38, 1.195]	[0.057, 0.041]
0.76	2.97	[1.44, 1.236]	[0.058, 0.044]
0.82	3.12	[1.499, 1.281]	[0.057, 0.0467]
0.96	3.20	[1.556, 1.327]	[0.368, 0.353]

Τελικά προκύπτει η ευθεία $y = 1.59 + 1.363x$

$Ερώτημα\ 4$

Εκτελώντας την παραπάνω διαδικασία για 100 εποχές στην Python παίρνω μια πολύ καλύτερη προσέγγιση $y = 1.32 + 2.07x$.

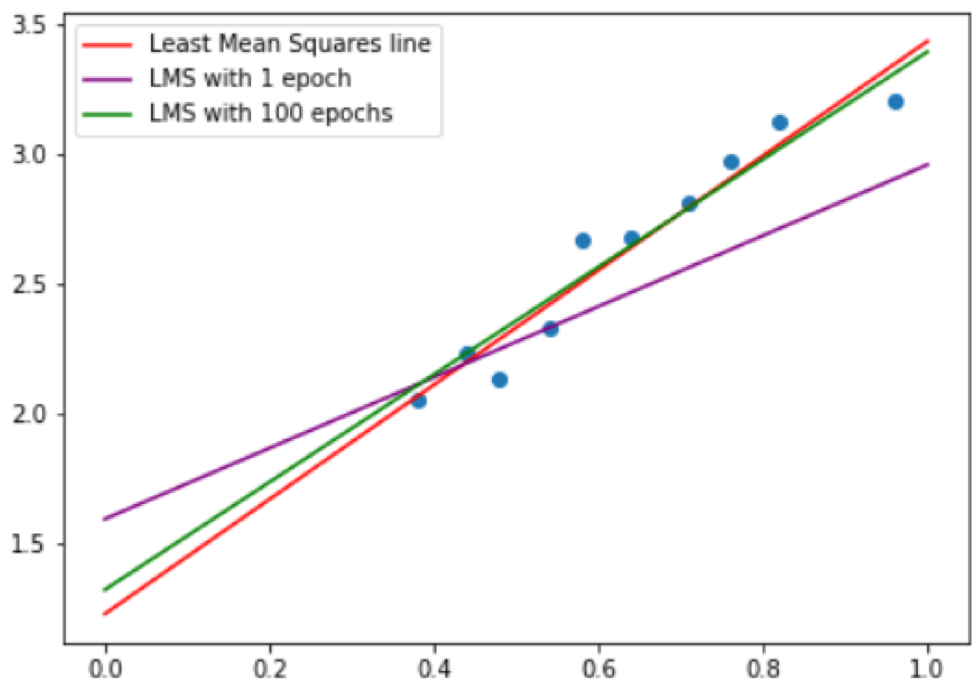
Παρακάτω δίνεται ο κώδικας

```python
def initialise_weights(size):
  return np.ones(size)

def update_weights(w, x, y, r = 0.05):
  pred = np.dot(w, x)
  diff = y - pred
  update = r * diff * x
  return np.add(w, update)

def mean_square_error(w, samples, values):
  error = 0
  for x, y in zip(samples, values):
    pred = np.dot(w, x)
    diff = y - pred
    error += diff ** 2
  return 0.5 * error
```

```python
def lms(samples, values, r=0.05, max_epoch=5):
  w = initialise_weights(samples.shape[1])
  w_new = initialise_weights(samples.shape[1])
  epoch = 0
  while epoch < max_epoch:
    for x, y in zip(samples, values):
      error = mean_square_error(w, samples, values)
      w_new = update_weights(w, x, y, r)
      w = w_new.copy()
    epoch += 1
  return w
```

Ερώτημα 5

Όλα τα παραπάνω αποτελέσματα αποτυπώνονται στην γραφική παράσταση



Άσκηση 1.7

Ερώτημα 1

Η κλάση ενός δείγματος αποφασίζεται με βάση το πρόσημο της διαφοράς των δύο κατανομών

$$g(x) = p(\omega_1|x) - p(\omega_2|x) = p(x|\omega_1)p(\omega_1) - p(x|\omega_2)p(\omega_2)$$

Άρα για $x \in [0, 1]$ ισχύει

$$g(x) = 2x - 1$$

Άρα το σημείο απόφασης είναι το

$$g(x) = 0 \Rightarrow x = \frac{1}{2}$$

Η πιθανότητα σφάλματος είναι η εξής

$$P(error) = \int_{R_1} p(\omega_2|x)dx + \int_{R_2} p(\omega_1|x)dx = \int_{\frac{1}{2}}^{1} 1 - xdx + \int_{0}^{\frac{1}{2}} xdx = \frac{1}{4}$$

Ερώτημα 2

Έστω ότι επιλέξαμε τα $x_1 \in \omega_1, x_2 \in \omega_2$ και $x \in \omega_1$. Η πιθανότητα σφάλματος του νέου ταξινομητή είναι η πιθανότητα το x να βρίσκεται πιο κοντά στο x_2 από ότι στο x_1. Άρα έχουμε την ανισότητα

$$|x - x_1| \leq |x - x_2| \Rightarrow$$

$$(x - x_1)^2 \leq (x - x_2)^2 \Rightarrow$$

$$(x_1 - x_2)(x_1 + x_2 - 2x) \geq 0 \Rightarrow$$

$$x_1 \geq x_2, x \leq \frac{x_1 + x_2}{2} \text{ ή } x_1 \leq x_2, x \geq \frac{x_1 + x_2}{2}$$

Άρα

$$P[error] = P[d_1 \geq d_2|x, x_1 \in \omega_1, x_2 \in \omega_2] =$$
$$= \int_0^1 p(x_1|\omega_1) \int_0^{x_1} p(x_2|\omega_2) \int_0^{\frac{x_1+x_2}{2}} p(x|\omega_1)dxdx_2dx_1$$
$$+ \int_0^1 p(x_1|\omega_1) \int_{x_1}^{1} p(x_2|\omega_2) \int_{\frac{x_1+x_2}{2}}^{1} p(x|\omega_1)dxdx_2dx_1$$

Άσκηση 1.8

Το $Q(\theta,\theta^0)$ υπολογίζεται ως εξής

$$Q(\theta,\theta^0) = \int [\sum lnp(x_k|\theta)] * p(x_{32}|\theta^0, x_{31}=1)d_{x_{32}}$$

$$= \ln p(x_1|\theta) + lnp(x_2|\theta) + \int ln\, p(x_3|\theta) * \frac{p(x_3|\theta^0)}{\int p(x_3'|\theta^0)dx_{32}'}dx_{32}$$

$$= \ln p(x_1|\theta) + lnp(x_2|\theta) + 2e^2 \int lnp(x_3|\theta) * p(x_3|\theta^0)dx_{32}$$

Θα πρέπει να ισχύει $\theta_2 \geq 5$ αλλιώς $Q(\theta,\theta^0) = -\infty$, οπότε

$$Q(\theta,\theta^0) = \ln p(x_1|\theta) + lnp(x_2|\theta) + 2e^2 \int_0^3 lnp(x_3|\theta) * p(x_3|\theta^0)dx_{32}$$

$$= -6\theta_1 - 3ln\theta_1\theta_2$$

Λόγω κανονικοποίησης της πρώτης κατανομής θα πρέπει

$$\int \frac{1}{\theta_1} e^{-\theta_1 x_1} dx_1 = 1 \Rightarrow \theta_1 = 1$$

Αντικαθιστώντας

$$Q(\theta,\theta^0) = -6 - 3ln\theta_2$$

Η οποία συνάρτηση μεγιστοποιείται για $\theta_2 = 5$ λόγω του περιορισμού και επειδή είναι φθίνουσα.

Άρα οι κατανομές πριν και μετά είναι οι εξής