Question 1

(a) Maximum Likelihood Estimate

Pr(X = 1) = 3 / 10 = 0.3, because {1} has 3 times

Pr(X = 4) = 1 / 10 = 0.1, because {4} has 1 time

(b) Laplace's Rule of Succession

Pr(X = 4) = 1 / 10 —> 1+π / 10+6π

Question 2

(a) Query x = {*Sunny, Hot, Normal, False*}

(a-1) k-NN (k-Nearest Neighbors algorithm), k = 3

Outlook		Temperature		Humidity		Windy		Play(Label)	Distance
0	Sunny	0	Hot	1	High	0	FALSE	-1	1
0	Sunny	0	Hot	1	High	1	TRUE	-1	2
1	Overcast	0	Hot	1	High	0	FALSE	+1	2
1	Rainy	1	Mild	1	High	0	FALSE	+1	3
1	Rainy	1	Cool	0	Normal	0	FALSE	+1	2
1	Rainy	1	Cool	0	Normal	1	TRUE	-1	3
1	Overcast	1	Cool	0	Normal	1	TRUE	+1	3
0	Sunny	1	Mild	1	High	0	FALSE	-1	2
0	Sunny	1	Cool	0	Normal	0	FALSE	+1	1
1	Rainy	1	Mild	0	Normal	0	FALSE	+1	2
0	Sunny	1	Mild	0	Normal	1	TRUE	+1	2
1	Overcast	1	Mild	1	High	1	TRUE	+1	4
1	Overcast	0	Hot	0	Normal	0	FALSE	+1	1
1	Rainy	1	Mild	1	High	1	TRUE	-1	4

'+1' × 2 and '-1' × 1. x = {*Sunny, Hot, Normal, False*}, Play = y = +1.

(a-2) Bayes’ rule

Let { *Sunny, Hot, Normal, False* } = { A1, A2, A3, A4 }

P(1 | A1, A2, A3, A4) = (P(1) * P(A1, A2, A3, A4 | 1)) / P(A1, A2, A3, A4) ······ ①

P(1) = 9/14

P(A1, A2, A3, A4 | 1) = P(A1 | 1) * P(A2 | 1) * P(A3 | 1) * P(A4 | 1)

= (2/9) * (2/9) * (6/9) * (6/9) = 16/729

P(A1, A2, A3, A4) = (5/14) * (4/14) * (7/14) * (8/14) = 10/73

① = ((9/14) * (16/729)) / (10/73) = 196/405 ≒ 0.484

as above, P(-1 | A1, A2, A3, A4)

= ((5/14) * (3/5) * (2/5) * (1/5) * (2/5)) / (10/73)

= 147/625 = 0.2352

∵ P(1 | A1, A2, A3, A4) = 0.484 > P(-1 | A1, A2, A3, A4) = 0.2352

∴ Play = y = +1

(b) Query x = {*Overcast, 98, Normal, True*}

(b-1) k-NN (k-Nearest Neighbors algorithm), k = 1

Outlook		Temperature		Humidity		Windy		Play(Label)	Distance
0	Sunny	0.1212	102	1	High	1	FALSE	-1	2.1212
0	Sunny	0.2121	105	1	High	0	TRUE	-1	1.2121
1	Overcast	0.0909	101	1	High	1	FALSE	+1	3.0909
1	Rainy	0.2424	90	1	High	1	FALSE	+1	3.2424
1	Rainy	0.9091	68	0	Normal	1	FALSE	+1	2.9091
1	Rainy	0.7879	72	0	Normal	0	TRUE	-1	1.7879
1	Overcast	1	65	0	Normal	0	TRUE	+1	2
0	Sunny	0.1212	94	1	High	1	FALSE	-1	2.1212
0	Sunny	0.7273	74	0	Normal	1	FALSE	+1	1.7273
1	Rainy	0.3636	86	0	Normal	1	FALSE	+1	2.3636
0	Sunny	0.303	88	0	Normal	0	TRUE	+1	0.303
1	Overcast	0.2424	90	1	High	0	TRUE	+1	2.2424
1	Overcast	0.0606	100	0	Normal	1	FALSE	+1	2.0606
1	Rainy	0.303	88	1	High	0	TRUE	-1	2.303

How do I compute the distance in 'Temperature' ?

Distance in Temperature (D in T)=

ROUND(ABS(T – 98) ÷ MAX((MAX(T) – 98), (98 – MIN(T))), 4)

Why don't I use k = 3, 5 or 7 in this case ?

We can see distance of the nearest point is 0.303, but 2nd and 3rd are more than 1.2121 (x4 long with 1st point to origin). They are too far to make the centroid close to the last 2 points.

'+1' × 1. x = {*Overcast, 98, Normal, True*}, Play = y = +1.

(b-2) Bayes' rule

$f_{T \mid P=1}(x) = (1 / (4*\sqrt{2\pi}))\ e\hat{}((-1/32)(x-85)^2)$

$f_{T \mid P=-1}(x) = (1 / (5*\sqrt{2\pi}))\ e\hat{}((-1/50)(x-92)^2)$

Suppose $f_{T \mid P=1}(98) = P(A2 \mid 1)$

$f_{T \mid P=-1}(98) = P(A2 \mid -1)$

P(1) P(1 | A1, A2, A3, A4)

$= (9/14) * (4/9) * (1 / (4*\sqrt{2\pi}))\ e\hat{}((-1/32)(98-85)^2) * (6/9) * (3/9)$

$= 3.22 * 10^{-5}$

P(-1) P(-1 | A1, A2, A3, A4)

$= (5/14) * (0/5) * (1 / (5*\sqrt{2\pi}))\ e\hat{}((-1/50)(98-92)^2) * (1/5) * (3/5)$

$= 0$

$\because P(1 \mid A1, A2, A3, A4) = 3.22 * 10^{-5} > P(-1 \mid A1, A2, A3, A4) = 0$

$\therefore$ Play = y = +1

Question 3 (Source code : https://github.com/grass0916/NTUST-1041-ML-HW01)

(a) Find the hypothesis by Dual Form Perceptron Algorithm by the training data A+ and A-.

(b) Generate 10,000 points in the box and plot the points for h(x) from (a).

Pseudocode

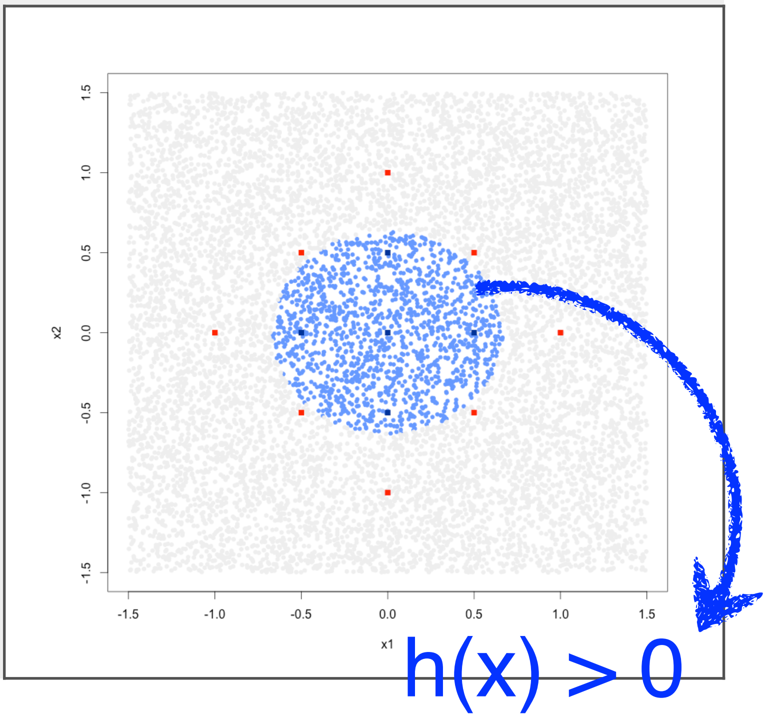
```
a = 0 ; b = 0 ; R = max{ ||xi||^2 }, i = 1 to l

while still error
  for i = 1 to l
    if yi * Σ ( aj * yj * <xi·xj>^2 ) + b <= 0, j = 1 to l
      ai <- ai + 1
      b  <- b  + yi * R^2
    end if
  end for
  return a, b to define h
end while until no error occurs then return a, b to define h
```

Hypothesis

```
h(x) = sign ( Σ ( ai * yi * < xi'·x >^2 ) + b ), i = 1 to l
```

	x1	x2	a	y
[1]	0.0	0.0	10	1
[2]	0.5	0.0	1	1
[3]	0.0	0.5	0	1
[4]	-0.5	0.0	0	1
[5]	0.0	-0.5	0	1
[6]	0.5	0.5	5	-1
[7]	0.5	-0.5	5	-1
[8]	-0.5	0.5	0	-1
[9]	-0.5	-0.5	0	-1
[10]	1.0	0.0	0	-1
[11]	0.0	1.0	0	-1
[12]	-1.0	0.0	0	-1
[13]	0.0	-1.0	0	-1

b = 1

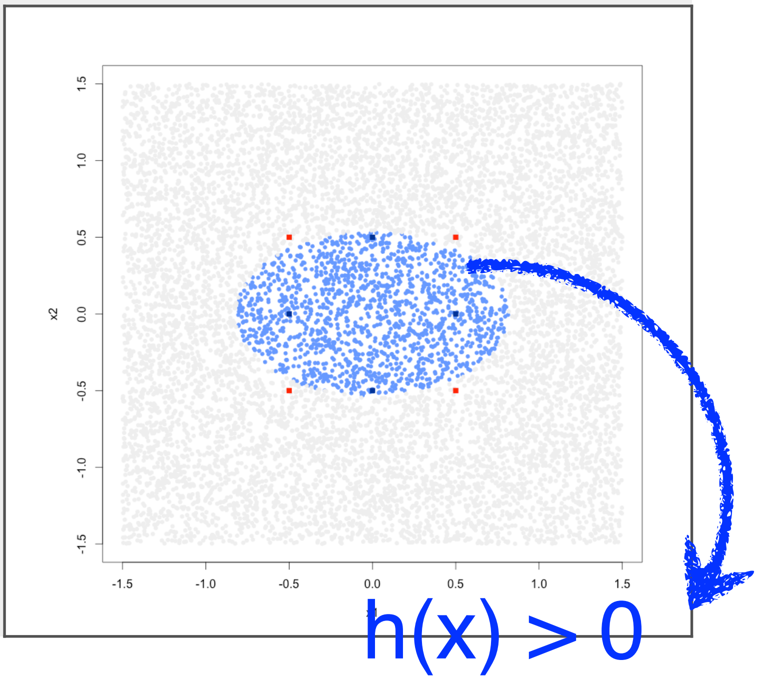


(c) Repeat (a) and (b) by using the training data B+ and B-.

Hypothesis

```
h(x) = sign ( Σ ( αi * yi * < xi'·x >^2 ) + b ), i = 1 to l
```

```
       x1    x2    α    y
[1]   0.5   0.0    9    1
[2]   0.0   0.5    5    1
[3]  -0.5   0.0    0    1
[4]   0.0  -0.5    0    1
[5]   0.5   0.5    6   -1
[6]   0.5  -0.5    6   -1
[7]  -0.5   0.5    0   -1
[8]  -0.5  -0.5    0   -1

 b = 0.5
```



(d) Mapping training data A+ and A- to $\phi : R^2 \rightarrow R^4$. And find hypothesis f(x) by Primal Perceptron Algorithm. $\phi(x) = [\ x1,\ x1^2,\ -x1x2,\ x2^2\ ]$

(e) Generate 10,000 points in the box and plot the points for f(x) from (d).

Pseudocode

```
b = 0 ; k = 0 ; w = 0 ; R = max{ ||xi||^2 }, i = 1 to l

while still error
  for i = 1 to l
    if yi * ( < w·φ(xi) > + b ) <= 0
      w <- w + 1 * yi * t(xi)
      b <- b  + yi * R^2
    end if
  end for
  k <- k + ifelse error then 1 or 0
end while until no error occurs then return w, b to define f
```

Hypothesis

```
f(x) = sign ( < w'·φ(xi) > + b ), i = 1 to l
```

	z1	z2	z3	z4	y
[1]	0.00	0.00	0.00	0.00	1
[2]	0.00	0.25	0.00	0.00	1
[3]	0.00	0.00	0.00	0.25	1
[4]	0.00	0.25	0.00	0.00	1
[5]	0.00	0.00	0.00	0.25	1
[6]	0.25	0.25	-0.25	0.25	-1
[7]	-0.25	0.25	0.25	0.25	-1
[8]	-0.25	0.25	0.25	0.25	-1
[9]	0.25	0.25	-0.25	0.25	-1
[10]	0.00	1.00	0.00	0.00	-1
[11]	0.00	0.00	0.00	1.00	-1
[12]	0.00	1.00	0.00	0.00	-1
[13]	0.00	0.00	0.00	1.00	-1

```
b = 1 ; k = 10 ;
w = [0.00, -2.25, 0.00, -2.50]
```