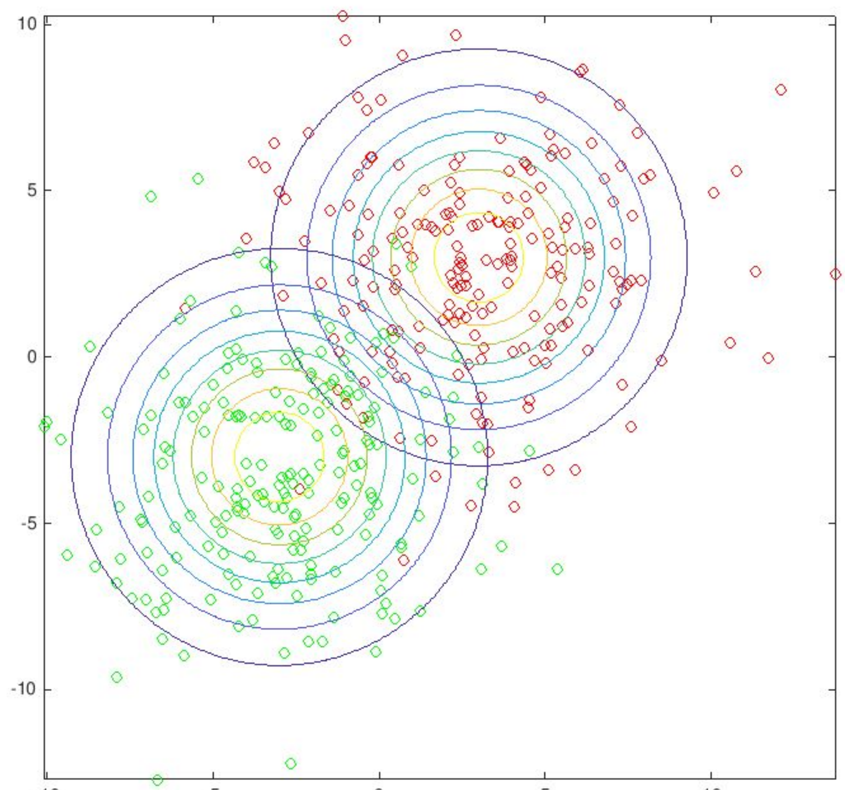
Hakkı Selim Öztürk
2013401027

25.12.18

EE 573 Pattern Recognition – Project 2 - Parameter Estimation

1. Generated samples and pdf contours



2. Estimating µi by considering µi as an unknown deterministic variable is Maximum Likelihood Estimation.

$$g1(x1, x2) \ = \ -\ 0.056 * x1\hat{}2 \ + \ 0.34 * x1 \ - \ 0.056 * x2\hat{}2 \ + \ 0.3 * x2 \ - \ 3.1$$
$$g2(x1, x2) \ = \ -\ 0.056 * x1\hat{}2 \ - \ 0.33 * x1 \ - \ 0.056 * x2\hat{}2 \ - \ 0.39 * x2 \ - \ 3.4$$

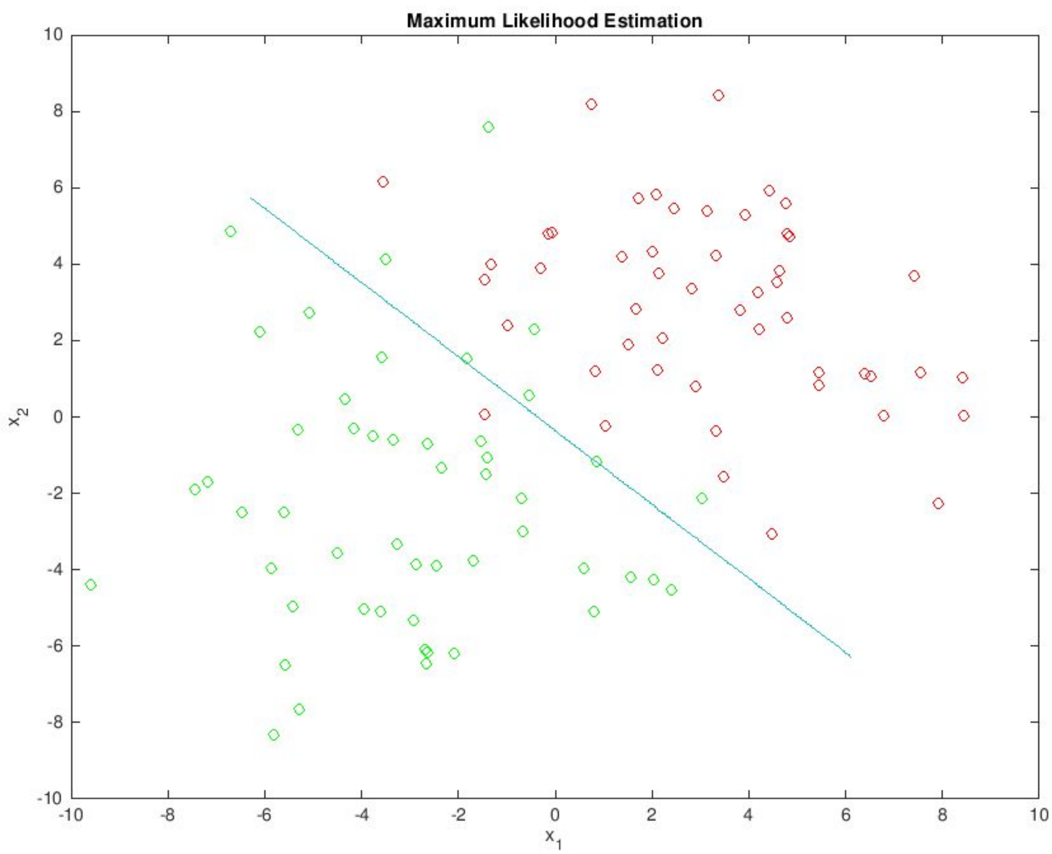
$$w \ = \ g1(x1, x2) \ - \ g2(x1, x2) \ \ = \ 0.67 * x1 \ + \ 0.7 * x2 \ + \ 0.25$$

Thus the decision rule is

$w1 \ \mathit{if} \ w > 0$

$w2 \ \mathit{otherwise}$

Which is



w1

precision	recall
0.8750	0.9800

w2

precision	recall
0.9773	0.8600

b. Considering μi as a random vector corresponds to the **Bayesian parameter estimation.**

$with\ \alpha:\ 2.9$

$\alpha 2:\ -3.1$

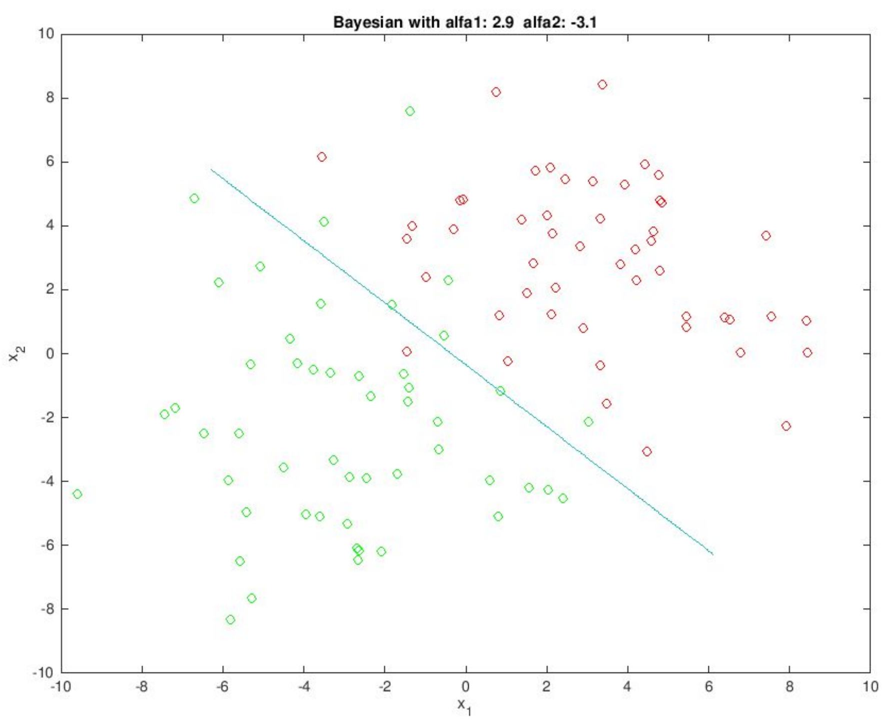
$Estimate\ for\ \mu 1\ =\ 3.0680,\ 2.7437$

$Estimate\ for\ \mu 2\ =-2.995,\ -3.5068$

$g1(x1,x2)\ =\ -\ 0.0556 * x1\hat{}2\ +\ 0.341 * x1\ -\ 0.0556 * x2\hat{}2\ +\ 0.305 * x2\ -\ 3.14$

$g2(x1,x2)\ =\ -\ 0.0556 * x1\hat{}2\ -\ 0.333 * x1\ -\ 0.0556 * x2\hat{}2\ -\ 0.39 * x2\ -\ 3.38$

$w\ =\ g1(x1,x2)\ -\ g2(x1,x2)\ \ = 0.674 * x1\ +\ 0.695 * x2\ +\ 0.241$



This is almost the same result with the first approach.

w1

precision	recall
0.8750	0.9800

w2

precision	recall
0.9773	0.8600

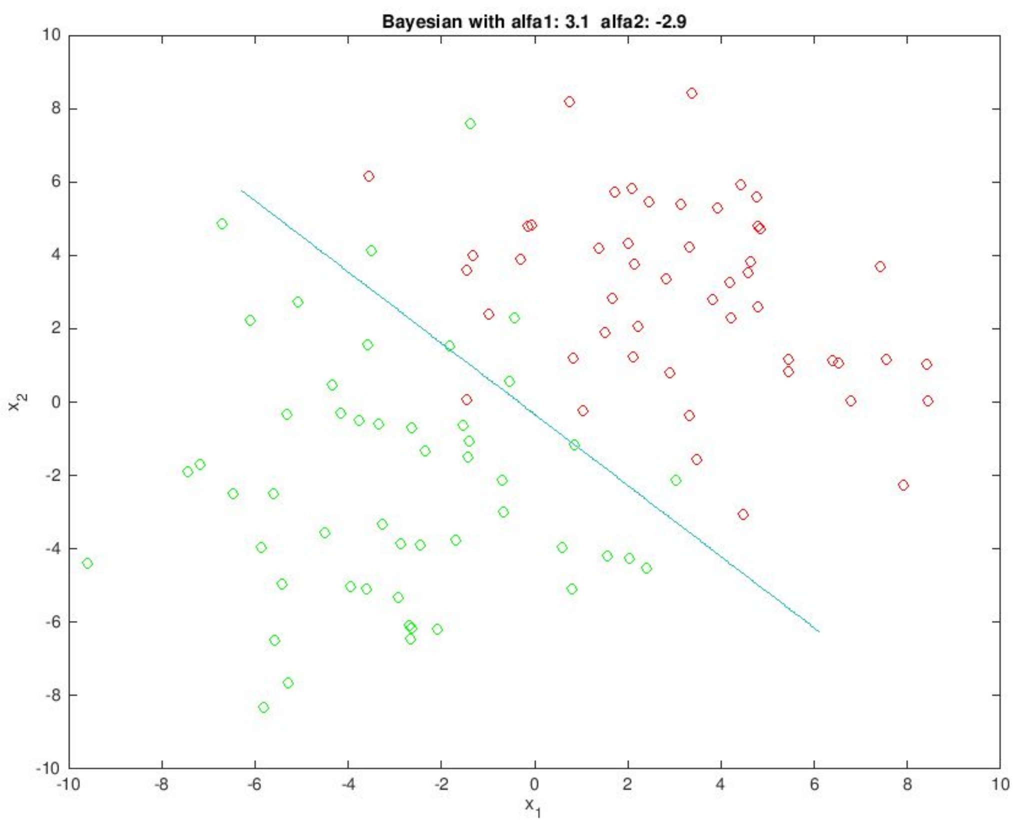
Precision and recall are exactly the same.

$with\ \alpha:\ 3.1$

$\alpha 2:\ -2.9$

$Estimate\ for\ \mu 1\ =\ 3.0767,\ 2.7523$

$Estimate\ for\ \mu 2\ =-2.9871,\ -3.4982$

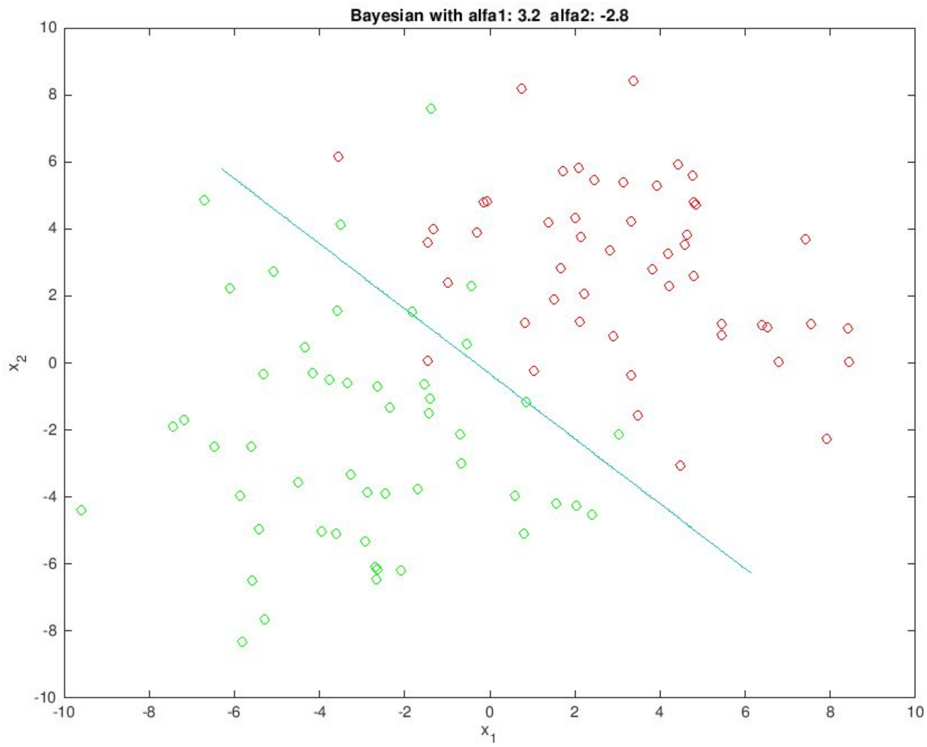


$with\ \alpha:\ 3.2$

$\alpha 2:\ -2.8$

$Estimate\ for\ \mu 1\ =\ 3.0810,\ 2.7566$

$Estimate\ for\ \mu 2\ =-2.9827,\ -3.4939$



All epsilon values leads to highly similar results.

Comparison

Maximum Likelihood Estimation is easier to compute because it uses differential calculus rather than integrals. It is also easier to interpret since it returns one model
$p(x|\theta)$ has the assumed parametric form in MLE.

Bayesian Estimation is harder to compute, because it needs multidimensional integration.
It is also harder to interpret because it returns weighted average of models

Can give better results since use more information about the problem (prior information)