Data Analysis and Data Mining
Lab Perceptron

1) Datasets

Five datasets for classification are given. Each dataset is stored using two text files: name_x and name_y. The first file contains the samples and the second one contains the labels. Display the data in "or", "xor", "iris" and display two samples of "face" (1 male, 1 female).

Dataset name	Description
or	The logical OR
xor	The logical XOR
face	A set of 90 images (60x60 pixel) of male and female faces [1]
sonar	A set of 208 samples, of 60 features each, representing sonar signals bounced off a metal cylinder or a roughly cylindrical rock [2]
iris	The Iris flower Fisher's dataset, reduced to 2 features (150 samples, three classes) [3]. Class 1 is linearly separable respect to Class2+Class3.

[1] M.Gray et al. "A Perceptron Reveals the Face of Sex", Neural Computation, Vol. 7, 1995.

[2] R.P.Gorman, T.J.Sejnowski " Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets" Neural Networks, Vol. 1, 1988.

[3] http://en.wikipedia.org/wiki/Iris_flower_data_set

2) Algorithm

Implement the Perceptron Learning Algorithm (PLA) and apply it to the linearly separable datasets.

Try different starting points and discuss the performance of the algorithm (e.g. how many iterations ? Is the number of iterations in agreement with the Perceptron Learning Theorem ?)

Sketch of the Perceptron Learning Algorithm:

```
read x,y
initialize w,b
err = n. of misclassified samples
i = 1
while err > 0
      f = w*x(i)+b
      if y(i)*f <= 0 then
            w := w + y(i)*x(i)
            b := b + y(i)
            err = n. of misclassified samples
      endif
      i := i + 1
      if (i > n. samples) i = 1
end while
```

3) Interpretation of results

Plot the linear separator (f(x) or f(x)=0), found by the PLA, and the corresponding samples of the "or and "iris" datasets.

Display the weights of the perceptron, which learned the "face" dataset, as a 60x60 image: what can be inferred from this image? (Note: consider the absolute value of the weights, discarding their sign)

4) Error estimation

Divide the "face" and "sonar" datasets in two halves (i.e. training and test sets) and compute the estimated error of the trained classifier. Try different starting points for the algorithm and discuss the results.