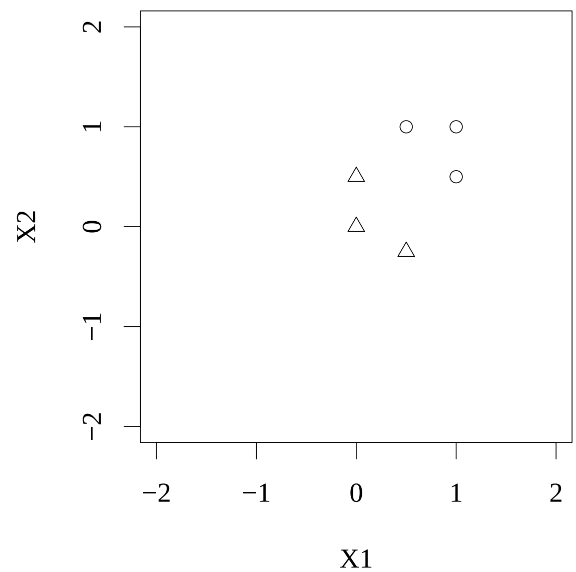
Introductory Applied Machine Learning, Tutorial Number 4

School of Informatics, University of Edinburgh, Instructors: Chris Williams, Oisin Mac Aodha, Hiroshi Shimodaira

September 2020

1. Consider a SVM with a linear kernel run on the following data set

x_1	x_2	y
0.5	1.00	1
1.0	0.50	1
1.0	1.00	1
0.0	0.50	2
0.5	-0.25	2
0.0	0.00	2



   (a) Using your intuition, what weight vector do you think will result from training an SVM on this data set?
   (b) Plot the data and the decision boundary of the weight vector you have chosen.
   (c) Which are the support vectors? What is the margin of this classifier?

2. You want to fit a mixture model with two Gaussians a and b to the following set of numbers:

`-9, -8, -7, -6, -5, 5, 5, 6, 6, 7, 7, 8, 8, 9, 9`

   (a) If you were to do the procedure manually (leveraging human intuition), what would be the means μ_a and μ_b of the two Gaussians, and the mixing parameters p_a and p_b?
   (b) Now you want to fit the mixture model in a completely automatic fashion, where you start with a random setting of all parameters and learn them using the EM algorithm. Suppose that the initial random setting of parameters is as follows: $\mu_a = -10$, $\mu_b = -20$, $\sigma_a^2 = \sigma_b^2 = 1$, and $p_a = p_b = 0.5$. Run the EM algorithm for a single iteration to determine the parameter values μ_a, μ_b, p_a and p_b (you do not need to compute the variances σ_a^2 and σ_b^2). For your convenience, the EM update equations involving Gaussian a are provided below, equations for Gaussian b are obtained by replacing a with b as appropriate:

$$
\begin{aligned}
p(x_i|a) &= \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(x_i-\mu_a)^2}{2\sigma_a^2}\right) \\
a_i &= P(a|x_i) = \frac{p(x_i|a)p_a}{p(x_i|a)p_a + p(x_i|b)p_b} \\
\mu_a &= \frac{a_1x_1 + a_2x_2 + \ldots + a_nx_n}{a_1 + a_2 + \ldots + a_n}
\end{aligned}
$$

$$\sigma_a^2 = \frac{a_1(x_1-\mu_a)^2 + \ldots + a_n(x_n-\mu_a)^2}{a_1+a_2+\ldots+a_n}$$

$$p_a = (a_1+a_2+\ldots+a_n)/n$$

(c) If you continue running EM, are the parameter values μ_1, μ_2, p_1 and p_2 guaranteed to converge to the values you computed in the first part of this question? Why or why not?