2. Consider four data points $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{x}_1 = (0,-4)$, $\mathbf{x}_2 = (1,0)$, $\mathbf{x}_3 = (0,5)$, $\mathbf{x}_4 = (-1,0)$ and their labels $y_1 = -1$, $y_2 = -1$, $y_3 = 1$,$y_4 = 1$.

a) Consider a perceptron that classifies these points.

i) Is there a set of labels for which this data cannot be shattered? [5 marks]

ii) Find the test error if the best hypothesis is $g = \mathbf{w} = (0,-2,+2)$ and the true target function is $f(\mathbf{x}) = (0,+4,0)$. [10 marks]

iii) Apply the perceptron learning algorithm to find the parameters of the linear predictor. [20 marks]

b) Consider using a non linear transformation $\phi(\mathbf{x}) = (x_1^2, x_2^2)$.

i) What is the break point of the linear classifier after such transformation? [5 marks]

ii) How does the predictor compare to the linear classifier before the transformation? [5 marks]

iii) Is $\phi(\mathbf{x})$ a good transformation to separate the data points $\mathbf{x}_1 \ldots, \mathbf{x}_4$? [5 marks]

iv) Is there a classifier that can correctly classify the points given by $\phi(\mathbf{x}) = (x_1^2, x_2^2)$? [5 marks]

c) Consider three hypothesis classes that you can use to find the best predictor ($\phi(\mathbf{x}) = \mathbf{x}$, $\phi(\mathbf{x}) = (x_1^4, x_2^4)$, $\phi(\mathbf{x}) = \mathbf{x}\mathbf{x}^\top$). What method would you use and how would you apply it to automatically find the best model?

[15 marks]

d) During training with a non-linear transformation of data samples you observe low training error but high validation error. Briefly outline possible reasons using learning curves to illustrate your points regarding the model complexity, amount of regularisation and the number of training samples. Propose possible actions to improve learning.

[30 marks]