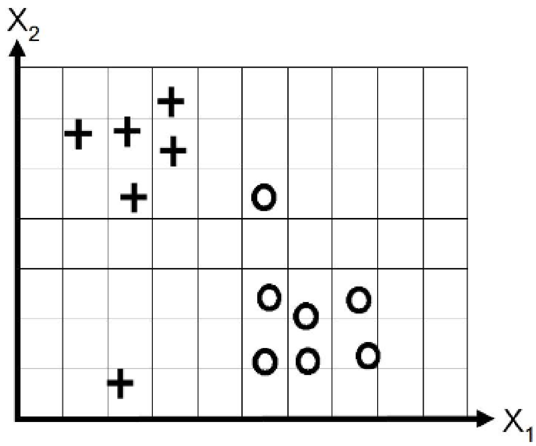
Quiz 8



(a) Consider the data in this figure, where we fit the model

$$p(y = 1 \mid x, w) = \sigma(w_0 + w_1 x_1 + w_2 x_2).$$

Suppose we fit the model by maximum likelihood, i.e., we minimize

$$J(w) = -\ell(w, \mathcal{D}_{\text{train}})$$

where $\ell(w, \mathcal{D}_{\text{train}})$ is the log likelihood on the training set. Sketch a possible decision boundary corresponding to $\hat{w}$. (Copy the figure first (a rough sketch is enough), and then superimpose your answer on your copy, since you will need multiple versions of this figure.) Is your answer (decision boundary) unique? How many classification errors does your method make on the training set?

(b) Now suppose we regularize only the w_0 parameter, i.e., we minimize

$$J_0(w) = -\ell(w, \mathcal{D}_{\text{train}}) + \lambda w_0^2$$

Suppose λ is a very large number, so we regularize w_0 all the way to 0, but all other parameters are unregularized. Sketch a possible decision boundary. How many classification errors does your method make on the training set? *Hint: consider the behavior of simple linear regression, $w_0 + w_1 x_1 + w_2 x_2$ when $x_1 = x_2 = 0$.*

(c) Now suppose we heavily regularize only the w_1 parameter, i.e., we minimize

$$J_1(w) = -\ell(w, \mathcal{D}_{\text{train}}) + \lambda w_1^2$$

Sketch a possible decision boundary. How many classification errors does your method make on the training set?

(d) Now suppose we heavily regularize only the w_2 parameter. Sketch a possible decision boundary. How many classification errors does your method make on the training set?