Exercise 3.6 [Cross-entropy error measure]

(a) More generally, if we are learning from ±1 data to predict a noisy target P(y|x) with candidate hypothesis h, show that the maximum likelihood method reduces to the task of finding h that minimizes

(b) For the case , argue that minimizing the in-sample error in part (a) is equivalent to minimizing the one in (3.9).

For two probability distributions {p, 1-p} and {q, 1-q} with binary outcomes, the cross-entropy (from information theory) is

The in-sample error in part (a) corresponds to a cross-entropy error measure on the data point with and .