

Machine Learning. Homework 1. Due March 6th (in class).

Problem 1. The feature space consists of three possible points (events) A, B, C , which occur with probability 0.1, 0.4, 0.5, respectively. For each event there are two possible labels $+1$ or -1 , which happen with probability 0.9, 0.3 and 0.8 respectively (that is, $P(1|A) = 0.9, P(1|B) = 0.3$, etc.). Determine the Bayes optimal classifier. What is the expected loss of the Bayes optimal classifier?

Problem 2. A probability distribution on the real line is a mixture of two classes $+1$ and -1 with density $N(1, 2)$ (normal distribution with mean 1 and variance 2) and $N(4, 1)$, with prior probabilities 0.4 and 0.6 respectively. What is the Bayes decision rule? Give an estimate for the Bayes risk.

Problem 3. Consider a k -NN classifier for a 2-class problem. What is its expected (classification) loss and how does it compare to the Bayes optimal, when $k = 3$, assuming you have sufficiently many data points? How does the empirical loss of 3-NN compare to the Bayes optimal? (Recall that the empirical loss of 1-NN is zero).

Problem 4. Generate 2000 points from two equally weighted spherical Gaussians $N(0, I), N((3, 0, \dots, 0), I)$ in \mathbb{R}^p , $p = 1, 11, 21, \dots, 101$ (note, you have to first flip a coin to decide from which Gaussian to sample), where I is the identity matrix and the centers of Gaussians are distance 3 apart. Implement 1-NN and 3-NN classifiers. Test the resulting classifier on a separately generated dataset with 1000 pts. Plot the error rate as a function of p . Observations?

Problem 5. What is the VC-dimension of the set of indicator functions of disks in \mathbb{R}^2 (i.e., functions which are 1 inside a circle -1 outside (but not the other way around!))? What about the indicator functions of rectangular boxes with sides parallel to the axes? You need to explain why.