

CS 4803A/8803A: Pattern Recognition

Problem Set 3

Date: Feb 20, 2003

Due: Feb 28th, 2003 - which is Friday before Spring Break.

But will be accepted until Monday 10th.

1. It is suspected that a certain four-sided die is unfair. By rolling the die many times, we wish to estimate the probability of each of the four outcomes. Let $\{P_i\}$, $i = 1, \dots, 4$ denote the unknown probabilities.

Suppose the die is rolled N times. If the probabilities to be estimated were known, then the probability of observing exactly y_1 1's, y_2 2's, \dots , and y_4 4's in the N rolls would be:

$$P(y_1, \dots, y_4 ; P_1, \dots, P_4) = \frac{N!}{y_1! y_2! \dots y_4!} P_1^{y_1} P_2^{y_2} \dots P_4^{y_4}.$$

This distribution is called the *multinomial distribution*.

Suppose exactly y_i occurrences of i are observed in the N rolls, where $i = 1, \dots, 4$. Find the maximum likelihood estimates $\{\hat{P}_i\}$ for the probabilities $\{P_i\}$, $i = 1, \dots, 4$ in terms of the y_i 's.

2. DHS Problem 3.2
3. DHS Problem 3.15
4. DHS Problem 3.11 (Symbolic math - but cool to see that the obvious Gaussian model is in fact the best Gaussian model for minimizing KL- divergence.)
5. **This one can be tricky. We'll do it in class but first do it yourself ...**

We return to the fishing scenario in Problem Set 2, Problem 7. By now you've constructed the optimal Bayesian classifier to decide if a fish is carp or bass, based on the uncertain class information you received from the two experts. You draw two fish from the pond which weigh 1.5 pounds and 5 pounds, respectively. If you applied your classifier independently to each fish, they would both be labeled bass. However, this may not seem right, and indeed it isn't (remember: only one expert is right about the pond). In this problem we will figure out what the optimal classification of the two fish should be.

- (a) The weight of a fish in the pond can be considered a random variable. If I draw two fish, the weights correspond to two random variables. Are these two variables independent, given your limited knowledge of the pond? Motivate your answer mathematically. (Remember: independence is not a property of the fish but of your knowledge about them.)

(b) There are four possible classifications of the two fish:

	$c_2 = \text{carp}$	$c_2 = \text{bass}$
$c_1 = \text{carp}$		
$c_1 = \text{bass}$		

where c_1 is the class of fish 1 (which weighs $w_1 = 1.5$ pounds) and c_2 is the class of fish 2 (which weighs $w_2 = 5$ pounds). Make a table like this and fill in the probability of each possible classification given the weights you observed. What is the most probable classification of the fish? What is the least probable classification of the fish?

(c) What does this result imply in general about performing pattern recognition when you are uncertain about the class distributions (e.g. because you estimated them from training data)?