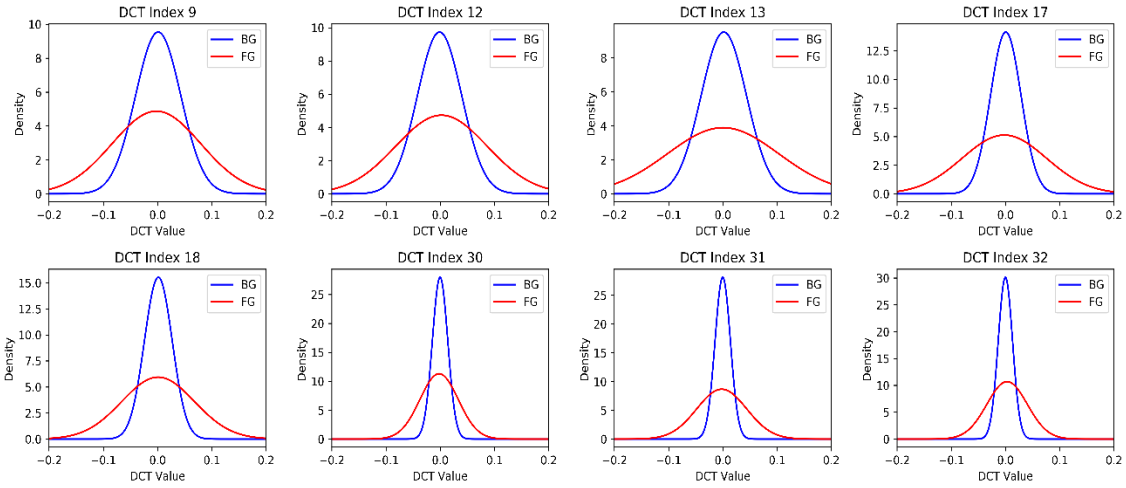
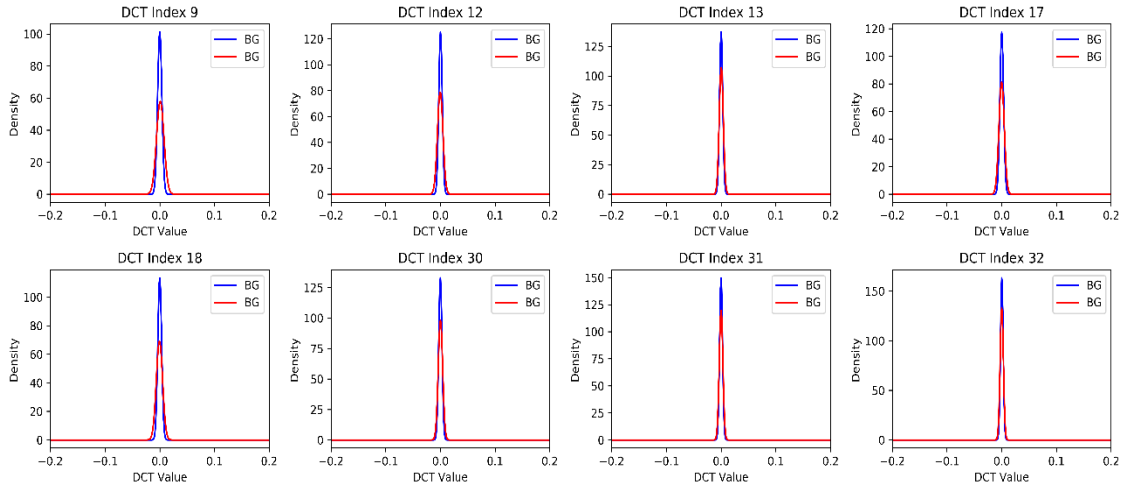
Richard Bell
ECE 271A
HW 2

a) The maximum likelihood estimates for the probability of a multinomial distribution is the number of successes over the total number of attempts. This means we count the total number of cheetah training samples and divide by the total number of samples and call this the probability of cheetah. We do the same thing to find the prior for background. This is exactly what we did last week with our intuitional estimator. Thus it turns out the intuitional estimator is the maximum likelihood estimator in this case.

b) The best features are those that behave very differently for one class then they do for the other. This difference allows us to determine which class we are currently dealing with high probability. When comparing Gaussian distributions, a difference can be caused by either having different means or different variances. In the plots shown below, the means are all very similar, so we cannot use them to tell the difference between classes. We are still ignoring the first index because it is too dependent on ambient lighting. Thus when selecting features, we should choose those features with the largest difference in variances. Features that share the same mean but different variances will more likely be produced by the class with the smaller variance. The best 8 and worst 8 features with this explanation in mind are shown below.



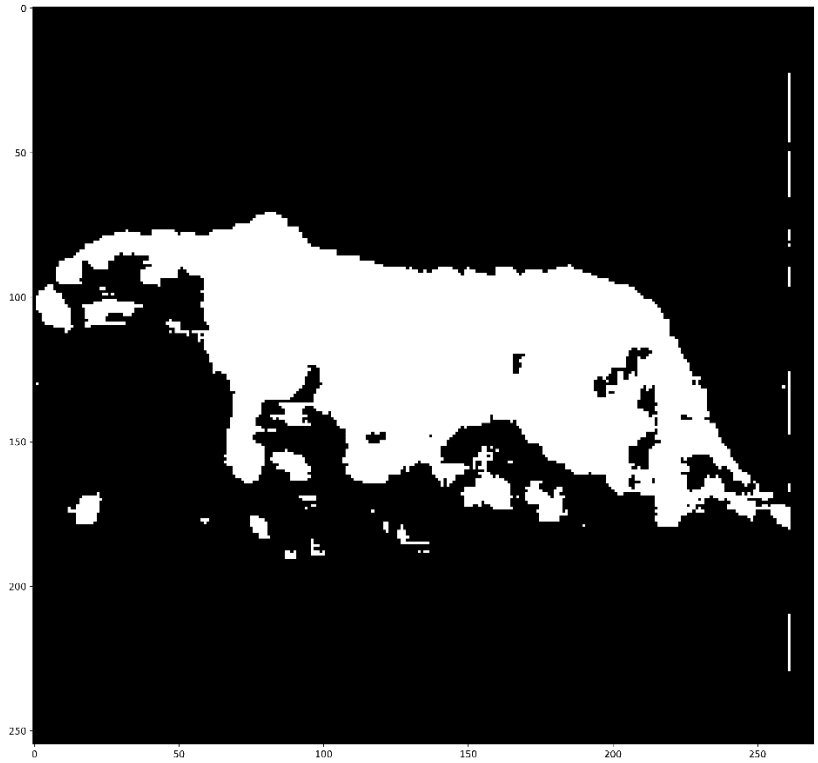
Good Features



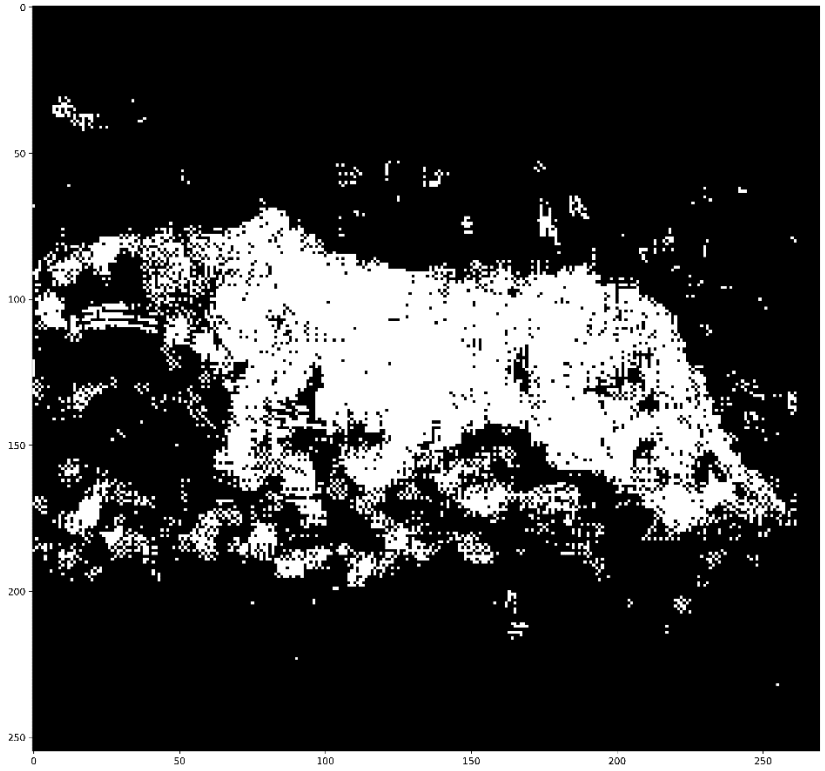
Bad Features

c) The Bayesian decision rule is given by: Choose FG if p(**x**|FG)p(FG) > p(**x**|BG)p(BG), where **x** is the vector of DCT values. We assume the components of **x** are independent random variables so that we can expand the joint density into the product of marginals to solve the decision rule numerically.

When all 64 features are used, the mask below is the result. The probability of error in this case is computed to be 6.66%.



When only the best 8 are used, the mask below is the result. The probability of error in this case is computed to be 10.18%.



When all 64 DCT coefficients are used we are taking advantage of all the information given to us when making the mask which results in a better mask. When only the best 8 features are used, we leave behind information that is useful which can be verified by the poorer quality of the second mask. In any situation, the algorithm that uses all the information provided to it will outperform others. The tradeoff is how much more complicated these algorithms may be. In practice one needs to weigh the gain in performance against the resources and time needed to compute them. The optimal algorithm is the one that meets all the specifications posed by the problem.