Aaron Dardik

CS567 Homework #3

* 1. For its left child has 150 examples in class A and 50 in class B. This implies its right child has 50 examples in class A and 150 in class B. For the left child has 0 examples in class A and 100 in class B. This implies its right child has 200 examples in class A and 100 in class B. Entropy is given by the following formula: For left) = left) = Still for the first tree we have that right) = right) = Thus, for we have that entropy of the left branch is and the entropy of the right branch is Now the entropy of is ½\*(entropy of left branch) + 1/2 \*(entropy of right branch). As the two branches are equal, the entropy of = entropy of left branch = entropy of right branch = 0.56. For we have that left) = 0, left) = 1. On the right branch we have right) = right) = For we have the entropy of the left branch is and the entropy of the right branch is For we have that its entropy is equal to ¼ \* (entropy of the left branch) + ¾ \* (entropy of the right branch) = Gini impurity is given by For on the left branch we have left)=left)= Therefore, for this branch the Gini impurity is As this tree displays symmetry, the Gini impurity of the right branch is also and as the Gini impurity of is ½ \*(impurity of the left branch) + ½ \*(impurity of the right branch) = For the Gini impurity of the left branch is For the right branch the Gini is The total Gini for We will now calculate each tree’s classification error. assigns the left branch to A and the right to B, so on the left side are misclassified. On the right branch are misclassified so the classification error for For the classification error on the left branch is 0, and on the right branch is so the total classification error of
  2. For we have entropy = 0.56, Gini impurity is 0.38 and classification error is 0.25. For entropy is 0.48, Gini impurity is 0.33 and classification error is 0.25. Based on classification error, the two trees are of the same quality, but using entropy and Gini impurity as metrics, causes to come out ahead. Therefore is of higher quality.

2.1) Let As is a convex function, its local minimum is also its global minimum. We will find the local, and therefore global minimum. Taking the derivative and setting it equal to 0, we get that We will perform a change of variables and let Thus, Rearranging, we get Now, as The second derivative test confirms this is indeed a minimum and as the local minimum of a convex function is the global minimum, we have shown that

2.2) We seek to prove that where the notation #(…) means “number of…” Similarly, and that

Now, Multiplying top and bottom by we see that the preceding equation simplifies to the following expression: . Remembering that so that and our preceding quotient will be reduced to This demonstrates the base case. We will now assume the inductive hypothesis, that Then, As we are summing over the cases where , we can see that the coefficient of in the numerator will never have a minus sign in front. This allows us to simplify the expression to . Now, using the inductive hypothesis and substituting back in for as well as recognizing that and that we can now simplify the quotient to and now by plugging in the values and simplifying further we reduce the quotient to the following form: as and the first sum on the right hand side of the preceding expression is Thus, our quotient reduces to and therefore the result holds for all t.

3.1) We would like to solve This is equivalent to finding The Lagrangian of this expression and the solution is found by the values that satisfy, Before differentiating, note that 0 is not a possible solution as it is not in the domain of the function due to being undefined. Now, taking Multiplying through by Now, as and using this to substitute back in to the expression for

3.2) Here we would like to solve the following problem: This is the same as finding We will now construct the Lagrangian. However, by the same reasoning as from problem 3.1 we see that the are all zero. So, we have Taking the derivative with respect to Simplifying, we have Note that this is equal to Again noting that so we can conclude that and therefore,

4.1) To solve this problem we will begin by solving to separate maximization problems. The first problem we will solve is Maximizing our function over is the same as minimizing -1 \* our function. We also require “shift” the greater than conditions to be We create the Lagrangian, with the requirement that implying Rearranging, we get that However, we see that together Therefore, And since and therefore, The second problem to solve is We put the equation into the required form for the KKT conditions as we did in the last step, and begin by taking the derivative of the Lagrangian with respect to and setting it equal to zero. Thus, Taking the derivative with respect to and setting it equal to zero, we get Our previous equation simplifies to By multiplying both sides of the equation by on the right, then dividing both sides by and then multiplying both sides on the left by we get

4.2) From problem 3.2 we see that Now solving and comparing to we have that and the result holds.

4.3) The loss function for K-means clustering can be written as If we consider a such that with the thus representing “how much,” or “the proportion of” of is in class We now write From this expression, if we set then we can recover the loss-function from K-means clustering as follows: the MLE for the Gaussian mixture model is found by maximizing and if for all *k* (i.e. all clusters have the same radius and are spherical) and is as described, then K-means clustering can be seen as a special case of the mixture model. The model parameters are therefore This shows that K-means clustering is a specific instance of the Gaussian mixture model where we consider the mixture weights to be equal and the variance to be spherical with each cluster having the same radius. The responsibility is given by which as variance goes to zero is equal to