Homework 1 - CMSC-25400

Patrick Collins

January 23, 2014

1 Proofs

1. Derive an algorithm that can learn any monotone conjunction over $0, 1^n$ with a mistake bound of n.

Algorithm 1 Addition for Monotone Conjunctions

$f \leftarrow \bigwedge_{n=1}^{n} x_n$
$i \leftarrow 1$
while True **do**
  predict $\hat{y}^t = f(x^t)$
  if $(\hat{y}^t == 1)$ and $(y^t == 0)$ **then**
    remove $\{x | x \in x^t \text{ if } x == 1\}$ from f.
  end if
  $t \leftarrow t + 1$
end while

Proof of mistake bound:

Note that (a) this algorithm will only ever produce false negatives, and (b) it removes at least one term from the concept for every false negative. Therefore, because there are only n terms total, it can make a maximum of n mistakes. ■

2. Prove that, for any concept class C and online algorithm A with finite mistake bound M, there is a conservative algorithm A' with mistake bound M on C.

Claim: Let A' be an algorithm that simulates the action of A on the input sequence S, which changes its hypothesis to the hypothesis currently held by A only when it makes a mistake. When such a mistake occurs, let A' remove every f that is inconsistent with the mistake from the concept class it is currently considering, $C(A')$. Then $M(A') \leq M(A)$, i.e. the mistake bound of A' is bounded from above by the mistake bound of A.

Lemma 1: After any sequence of inputs, $C(A)$ contains at least as many incorrect hypotheses as $C(A')$, and both contain the true hypothesis.

Proof: Suppose, for contradiction, that there exists some hypothesis $f \in C(A), f \notin C(A')$. Then either f is consistent with the input that both algorithms have processed so far, or it is not. If it is not consistent, then $C(A)$ has more incorrect hypotheses than $C(A')$, which is what we wanted to show. If f is consistent with the input processed so far, then A' would have never removed it from $C(A')$, a contradiction.

Alternatively, suppose, for contradiction, that there is some $f \notin C(A), f \in C(A')$. A' removes inconsistent hypotheses from its concept class, so f must be consistent with the input seen so far. But then f could be the correct hypothesis, and if A had removed the correct hypothesis from its concept class, then it would make infinitely many mistakes, a contradiction.

Proof of main claim: The mistakes that any algorithm makes is bounded from above by the number of incorrect hypotheses in the concept class its considering, so, the fact that $C(A') \leq C(A)$ over any sequence of inputs by lemma 1 implies that $M(A') \leq M(A)$ for any sequence of inputs, which is what we wanted to show. Hence A' has a mistake bound of M over C.
$\blacksquare$

3. Prove that $cost_{IC} = 2cost_{avg^2}$

   Beginning with the simple case of a single cluster center:

$$
\begin{aligned}
cost_{IC}(C) &= \frac{1}{|C|}\sum_x \sum_{x'} (x - x')^2 \\
&= \frac{1}{|C|}\sum_x \sum_{x'} ||x - x'|| \\
&= \frac{1}{|C|}\sum_x \sum_{x'} (x^T - x'^T)(x - x') \\
&= \frac{1}{|C|}\sum_x \sum_{x'} (x^T x - x^T x' - x'^T x + x'^T x') \\
&= \frac{1}{|C|}\sum_x \sum_{x'} x^T x - \frac{1}{|C|}\sum_x \sum_{x'} x^T x' - \frac{1}{|C|}\sum_x \sum_{x'} x'^T x + \frac{1}{|C|}\sum_x \sum_{x'} x'^T x'
\end{aligned}
$$

Noting that:

$$\sum_x \sum_{x'} x^T x = |C| \sum_x x^T x$$
$$\sum_x \sum_{x'} x'^T x' = |C| \sum_{x'} x'^T x'$$
$$\sum_{x'} x'^T x' = \sum_x x^T x$$

and

$$\sum_x \sum_{x'} x^T x' = \sum_x \sum_{x'} x'^T x$$

We can collect like terms:

$$\begin{aligned} \frac{1}{|C|} \sum_x \sum_{x'} (x^T x - x^T x' - x'^T x + x'^T x') &= 2 \sum_x x^T x - \frac{2}{|C|} \sum_x \sum_{x'} x^T x' \\ &= 2(\sum_x x^T x - \frac{1}{|C|} \sum_x x^T \sum_x x) \end{aligned}$$

Then, we begin to work from the other side of the equation. Noting that

$$m = \frac{\sum_{x'} x'}{|C|}$$

We have:

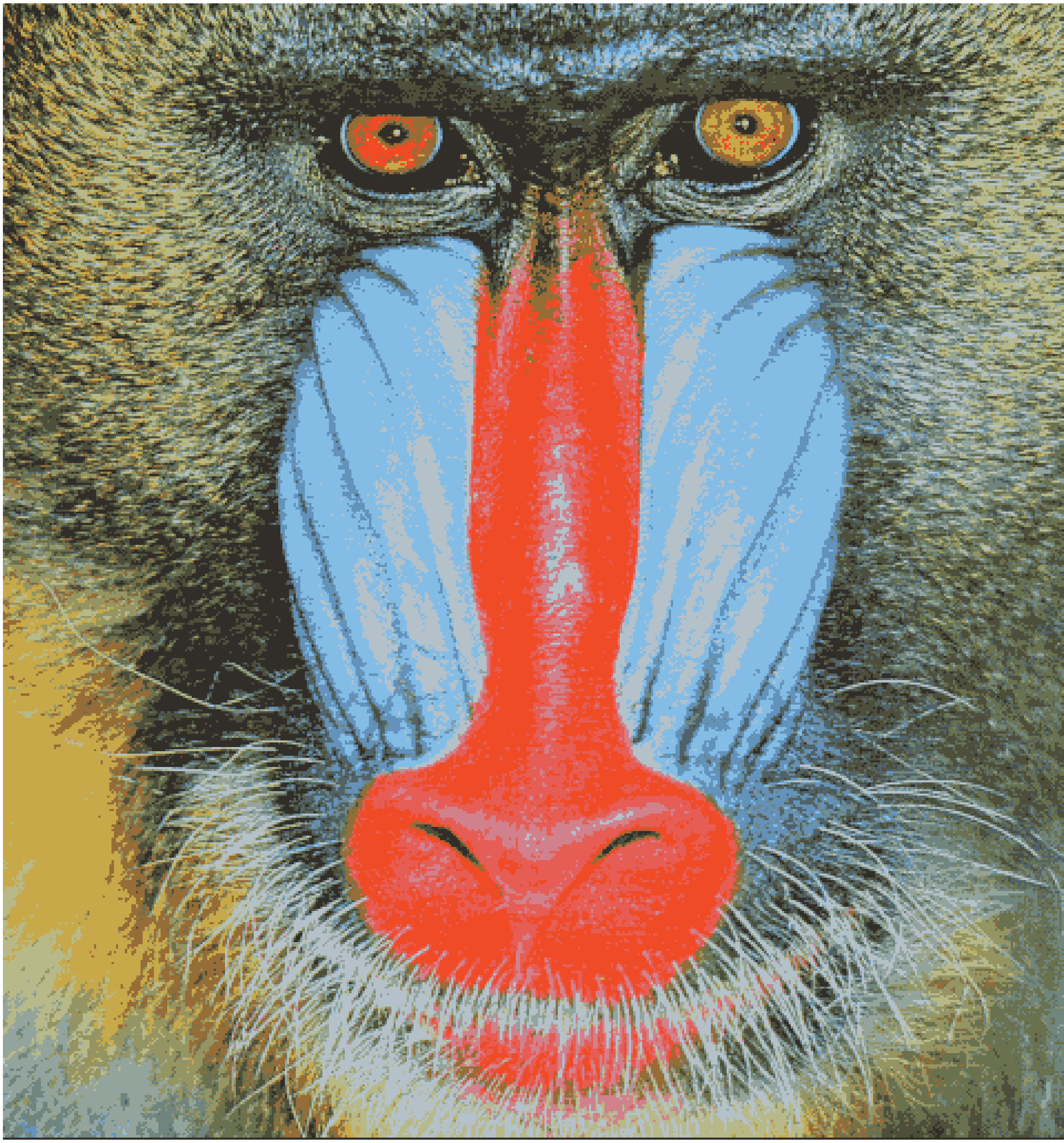
$$
\begin{aligned}
cost_{avg^2}(C) &= \sum_x d(x, \frac{\sum\limits_{x'} x'}{|C|})^2 \\
&= \sum_x ||x - \frac{\sum\limits_{x'} x'}{|C|}||^2 \\
&= \sum_x (x - \frac{\sum\limits_{x'} x'}{|C|})^T (x - \frac{\sum\limits_{x'} x'}{|C|}) \\
&= \sum_x (x^T x - x^T \frac{\sum\limits_{x'} x'}{|C|} - \frac{\sum\limits_{x'} x'^T}{|C|} x + \frac{\sum\limits_{x'} x'^T}{|C|} \frac{\sum\limits_{x'} x'}{|C|}) \\
&= \sum_x x^T x - \frac{1}{|C|} \sum_x x^T \sum_{x'} x' - \frac{1}{|C|} \sum_{x'} x'^T \sum_x x + \frac{1}{|C|^2} \sum_x x (\sum_{x'} x'^T \sum_{x'} x') \\
&= \sum_x x^T x - \frac{2}{|C|} \sum_x x^T \sum_{x'} x' + \frac{1}{|C|} \sum_{x'} x'^T \sum_{x'} x' \\
&= \sum_x x^T x - \frac{1}{|C|} \sum_x x^T \sum_x x
\end{aligned}
$$

Noting that the both the intra-cluster distance and the k-means cost within a single cluster do not depend on how many other clusters there are, we can simply add back in the summation on k, and get:

$$
\begin{aligned}
cost_{IC}(C_0 \cup C_1, \ldots, \cup C_k) &= \sum_i^k 2(\sum_{x \in C_i} x^T x - \frac{1}{|C_i|} \sum_{x \in C_i} x^T \sum_{x \in C_i} x) \\
&= 2 \sum_i^k (\sum_{x \in C_i} x^T x - \frac{1}{|C_i|} \sum_{x \in C_i} x^T \sum_{x \in C_i} x) \\
cost_{avg^2}(C_0 \cup C_1, \ldots, \cup C_k) &= \sum_i^k (\sum_x x^T x - \frac{1}{|x \in C_i|} \sum_x x^T \sum_x x)
\end{aligned}
$$

■

4. Results from applying KMeans:



The contrast in the resulting, compressed image is higher, reflecting the fact that the number of colors in the pictures has been halved. It also appears to be somewhat more washed-out than the original.

This compression reduces the total number of colors to 16, meaning that each pixel can be represented with three 4-bit numbers rather than 8-bit ones, resulting in a compression ratio of approximately 50% for large images, because the storage requirement for each pixel has been halved.