**CSC411 Homework 2**

**Question 1: Information Theory**

**Question 1a: Proof of Non-Negative Entropy**

Definition of entropy

Note that . Therefore, the above summation effectively ignores . Also note that all probabilities belong in the inclusive range .

Taking the log of the above:

Since is positive, multiplying by means that the range stays the same:

Since all in the summation have the same properties as the generic which was just analyzed:

Negating the summation to obtain the definition of entropy

Therefore, is non-negative.

**Question 1b: Proof of Non-Negative KL-Divergence**

Definition of KL-Divergence

Since

Note that the above is an expectation

Jensen’s inequality for **concave** function

Since is **concave** over the positive real numbers and is positive, Jensen’s inequality can be applied

Expanding the second expectation

Since the sum of probabilities over the entire set of events is 1

Rewriting the expectation in the inequality as negative KD-divergence

Negating the inequality

Therefore, KD-divergence is non-negative.

**Question 1c: KL-Divergence / Information Gain Equivalence**

Definition of KL-Divergence with and . Call this for brevity.

Applying Bayes’ theorem to the above:

Logarithm of division can be expressed as subtraction of logarithms

Substituting first term with definition of conditional entropy

Switching order of summation, and bringing outside the summation over

Applying the definition of marginal distribution

**Question 1c continued…**

Substituting second term with entropy of

The above is equivalent to information gain of given

**Question 2: Benefit of Averaging**

**Jensen’s inequality definition:** <https://en.wikipedia.org/wiki/Jensen%27s_inequality#Finite_form>

Denote the following for brevity (note that is fixed)

Jensen’s inequality for **convex** function , values and weights

If all weights , Jensen’s inequality for **convex** functions can be applied to averages of **convex** functions

Note that quadratic with respect to , so it is convex . Applying Jensen’s inequality for convex functions

Substituting definition of

Substituting definition of

Recalling definition of  , and cancelling the 2’s

**Question 3: AdaBoost**

Notes:

* Log refers to the natural logarithm, i.e.
* To avoid confusion with the training data (denoted as ), analysis will be conducted using as the iteration number.

Denote the following for brevity:

Rewriting the given definitions in terms of the above

**Want to Show:**

Denote the following sets

**Question 3 continued…**

Both and take values in the domain . Therefore, within each set and , the product is identical for all or all

|  |  |
| --- | --- |
|  |  |
| **Proof** | |
|  |  |
|  |  |
|  |  |
|  |  |

The above result can be used to create consistent definitions of for and

|  |  |
| --- | --- |
|  |  |
| **Proof** | |
|  |  |
|  |  |
| (Recalling definition of ) | (Negation of logarithm is inversion of operand) |
|  |  |
|  |  |
|  |  |

**Question 3, continued…**

Summation over the whole set is equal to adding the summations over and

Summation over all weights where is equal to summation over all weights for which

Using the previous two results to rewrite

Substituting the results for

Moving the radicals of outside the summations

Rewriting in terms of and

For brevity, let

Rewriting the fractions of that appear in using the above

𝑟\_𝑖𝑒𝑠 ted to create a summations(