The University of Texas at Dallas

CS 6375.001

Machine Learning

Assignment 2:

K-Means Unsupervised Learning Algorithm

Student:
Reg Gonzalez
rdg170330

Instructor:
Anurag Nagar

Due Date: April 28, 2024

Assignment 2 - Theoretical Part

Question #1.

Given:

$E_{agg}(x) = E[\{\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)\}^2]$

$E_{avg}(x) = \frac{1}{M}\sum_{i=1}^{M} E(\epsilon_i(x)^2)$

Assumptions:

(1) Each of the errors has a 0 mean (i.e., $E(\epsilon_i(x)) = 0\, for\ all\ i$)

(2) Errors are uncorrelated (i.e., $E(\epsilon_i(x)\epsilon_j(x)) = 0\, for\ all\ i \neq j$

Want to prove: $E_{agg} = \frac{1}{M}E_{avg}$.

First, let's expand $E_{agg}(x)$:

$E_{agg}(x) = E[\{\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)\}^2] = E[(\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)) * (\frac{1}{M}\sum_{j=1}^{M} \epsilon_j(x))]$

According to assumption (2), we only take into account the instance where $i = j$, since the error is 0 when $i \neq j$:

$E_{agg}(x) = E[(\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)) * (\frac{1}{M}\sum_{j=1}^{M} \epsilon_j(x))] = E[\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \epsilon_i(x)\epsilon_j(x)]$

Since we're assuming that $i = j$, we can rewrite the simplified equation above to be:

$E_{agg}(x) = E[\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \epsilon_i(x)\epsilon_j(x)] = E[\frac{1}{M^2}\sum_{i=1}^{M} \epsilon_i(x)^2] = \frac{1}{M^2}\sum_{i=1}^{M} E[\epsilon_i(x)^2]$

Assignment 2 - Theoretical Part

Remember that we're trying to prove that $E_{agg} = \frac{1}{M}E_{avg}$.

From above, we got $E_{agg}(x) = \frac{1}{M^2}\sum_{i=1}^{M} E[\epsilon_i(x)^2]$, and from the given equations, we know that $E_{avg}(x) = \frac{1}{M}\sum_{i=1}^{M} E(\epsilon_i(x)^2)$. If we multiply $E_{avg}(x)$ by $\frac{1}{M}$, we get $\frac{1}{M^2}\sum_{i=1}^{M} E(\epsilon_i(x)^2)$, which is exactly what we proved E_{agg} is.

Assignment 2 - Theoretical Part

Question #2.

<u>*Given:*</u>

$E_{agg}(x) = E[\{\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)\}^2]$

$E_{avg}(x) = \frac{1}{M}\sum_{i=1}^{M} E(\epsilon_i(x)^2)$

<u>*Assumptions:*</u>

(1) Each of the errors has a 0 mean (i.e., $E(\epsilon_i(x)) = 0 \; for \; all \; i$)

(2) Jensen's inequality (i.e., $f(\sum_{i=1}^{M} \lambda_i x_i) \leq \sum_{i=1}^{M} \lambda_i f(x_i)$)

<u>*Want to prove:*</u> $E_{agg} \leq E_{avg}$.

First, let's expand $E_{agg}(x)$:

$E_{agg}(x) = E[\{\frac{1}{M}\sum_{i=1}^{M} \epsilon_i(x)\}^2] = E[\frac{1}{M^2}(\sum_{i=1}^{M} \epsilon_i(x))^2] = \frac{1}{M^2}E[(\sum_{i=1}^{M} \epsilon_i(x))^2]$

Next, expand $E_{avg}(x)$:

$E_{avg}(x) = \frac{1}{M}\sum_{i=1}^{M} E(\epsilon_i(x)^2) = \frac{1}{M}\sum_{i=1}^{M} E((f(x) - h_i(x))^2)$

We can replace the $\epsilon_i(x)$ term with $f(x) - h_i(x)$ because that was given to us in the problem description as well: $\epsilon_i(x) = f(x) - h_i(x)$.

Likewise, we can also apply this fact to the equation we got for $E_{agg}(x)$ above:

$E_{agg}(x) = \frac{1}{M^2}E[(\sum_{i=1}^{M} \epsilon_i(x))^2] = \frac{1}{M^2}E[(\sum_{i=1}^{M} f(x) - h_i(x))^2]$

Assignment 2 - Theoretical Part

Now we have the two equations:

$$E_{avg}(x) = \frac{1}{M}\sum_{i=1}^{M} E((f(x) - h_i(x))^2)$$

$$E_{agg}(x) = \frac{1}{M^2}E[(\sum_{i=1}^{M} f(x) - h_i(x))^2]$$

Remember that one of our assumptions is Jensen's inequality, which states:

$$f(\sum_{i=1}^{M} \lambda_i x_i) \leq \sum_{i=1}^{M} \lambda_i f(x_i)$$

And we want to prove that $E_{agg} \leq E_{avg}$.

Let's rewrite $E_{agg} \leq E_{avg}$ in terms of the equations we got above:

$$\frac{1}{M^2}E[(\sum_{i=1}^{M} f(x) - h_i(x))^2] \leq \frac{1}{M}\sum_{i=1}^{M} E((f(x) - h_i(x))^2)$$

Not only does this equation hold true, but it also matches our assumption involving Jensen's inequality.

Reg Gonzalez CS 6375.001

Assignment 2 - Theoretical Part

Question #3.

Given:

- Final hypothesis for a Boolean classification problem at the end of T iterations: $H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$
- Weight for the point i at step t+1: $D_{t+1}(i) = \frac{D_t(i)}{Z_t} * e^{-\alpha_t h_t(i) y(i)}$
- At step 1, points have equal weights: $D_1 = \frac{1}{N}$
- Total error of h_t: $\varepsilon_t = \frac{1}{2} - \gamma_t$

Want to prove: The overall training error, at the end of T steps, is bounded by:
$exp(-2\sum_{i=1}^{T} \gamma_i^2)$

(1)

$D_{t+1}(i) = \frac{D_t(i)}{Z_t} * e^{-\alpha_t h_t(i) y(i)} \rightarrow$

Can be rewritten as $D_{t+1}(i) = \frac{D_t(i)}{Z_t} * e^{-\alpha_t h_t(x_i) y(i)} \rightarrow$

Can be expanded to $D_{t+1}(i) = [\frac{e^{-\alpha_1 h_1(i) y(i)}}{Z_1}] * [\frac{e^{-\alpha_2 h_2(i) y(i)}}{Z_2}] * \ldots * [\frac{e^{-\alpha_T h_T(i) y(i)}}{Z_T}] \rightarrow$

Can be rewritten as $D_{t+1}(i) = (\frac{1}{N})(\frac{e^{-y(i) * f(x_i)}}{\prod_{i=1}^{T} Z_T})$, where $f(x_i) = \sum_{i=1}^{T} \alpha_t h_t(x_i)$

(2)

The training error (H) is defined as: $\frac{1}{N}\sum_i \{1\ if\ y_i \neq H(x_i)\ or\ else\ 0\} \rightarrow$

Can be rewritten as $\frac{1}{N}\sum_i \{1\ if\ y_i * f(x_i) \leq 0\ or\ else\ 0 \rightarrow$

$\leq \frac{1}{N}\sum_i e^{(-y_i * f(x_i))}$, where $f(x_i) = \sum_{i=1}^{T} \alpha_t h_t(x_i) \rightarrow$

Assignment 2 - Theoretical Part

$= \sum_{i}^{T} D_{T+1}(i) * \prod_{t}^{T} Z_t = \prod_{t} Z_t$.

(3)

From given equation: $D_{t+1}(i) = \frac{D_t(i)}{Z_t} * e^{-\alpha_t h_t(i) y(i)} \rightarrow$

Rewrite to compute Z_t: $Z_t = \sum_{i} D_t(i) * \{e^{\alpha_t} \; if \; h_t(x_i) \neq y_i \; or \; e^{-\alpha_t} \; if \; h_t(x_i) = y_i\} \rightarrow$

Can be rewritten as $Z_t = \sum_{i;\, if \, h_t(x_i) \neq y_i} D_t(i) * e^{\alpha_t} + \sum_{i;\, if \, h_t(x_i) = y_i} D_t(i) * e^{-\alpha_t} \rightarrow$

Can be rewritten as $Z_t = (e^{\alpha_t})(\sum_{i;\, if \, h_t(x_i) \neq y_i} D_t(i)) + (e^{-\alpha_t})(\sum_{i;\, if \, h_t(x_i) = y_i} D_t(i)) \rightarrow$

Can be rewritten as $Z_t = e^{\alpha_t}\varepsilon_t + e^{-\alpha_t}(1 - \varepsilon_t) \rightarrow$

$= 2\sqrt{(1 - \varepsilon_t)\varepsilon_t} = \sqrt{1 - 4\gamma_t^2} \rightarrow$

$\leq e^{-2\gamma_t^2}$.

The result from (2) was $\prod_{t} Z_t$. If we combine this with the result from (3), $Z_t \leq e^{-2\gamma_t^2}$, we can see that this proves the claim that the training error is bounded by $exp(-2\sum_{i=1}^{T} \gamma_i^2)$.