Nam Nguyen

npn190000

1.

Consider $E\left[x_1 + x_2\right] = E\left[x_1\right] + E\left[x_2\right] = \sum_{i=1}^{2} E\left[x_i\right]$

The error using the aggregated model is defined as:

$E_{agg}(x) = E\left[\left\{\frac{1}{M}\sum_{i=1}^{M} h_i(x) - f(x)\right\}^2\right]$ *with M different models* $- h_1, h_2, \ldots, h_M.$

Prove: $E_{agg} = \frac{1}{M}E_{avg}$

$$
\begin{aligned}
E_{agg}(x) &= E\left[\left\{\frac{1}{M}\sum_{i=1}^{M} h_i(x) - f(x)\right\}^2\right] \\
&= E\left[\frac{1}{M^2}\left[\sum_{i=1}^{M} h_i(x) - f(x)\right]^2\right] \\
&= \frac{1}{M^2}.E\left[\sum_{i=1}^{M} h_i(x) - f(x)\right]^2 \\
&= \frac{1}{M^2}.E\left[\sum_{i=1}^{M} -\left(f(x) - h_i(x)\right)\right]^2
\end{aligned}
$$

The error for each of the models would be described as:

$$\varepsilon_i(x) = f(x) - h_i(x)$$

$$E_{agg}(x) = \frac{1}{M^2}.E\left[\sum_{i=1}^{M} -\left(\varepsilon_i(x)\right)\right]^2 = \frac{1}{M^2}.E\left[\sum_{i=1}^{M}\left(\varepsilon_i(x)\right)\right]^2$$

The average value of the expected squared error for each of the models acting individually is defined as:

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

$$E_{agg}(x) = \frac{1}{M}.E_{avg}\ (Proved)$$

2.

In question 1, we had assumed that each of the errors are uncorrelated i.e.
$E(\varepsilon_i(x)\varepsilon_j(x)) = 0\ for\ all\ i \neq j$

This is not really true, as the models are created using bootstrap samples and have correlation with each other. Now, let's remove that assumption. Show that using Jensen's inequality, it is still possible to prove that: $E_{agg} \leq E_{avg}$

Jensen's inequality states that for any convex function f

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

$$E[x] = \sum_{i=1}^{M} \lambda_i x_i \,; \quad E[f(x)] = \sum_{i=1}^{M} \lambda_i f\left(x_i\right)$$

We have simple case convex function f:

$$E[f(x)] = \lambda_1 f\left(x_1\right) + \lambda_2 f\left(x_2\right) \geq f\left(\lambda_1 x_1 + \lambda_2 x_2\right) = f(E[x])$$

Consider convex function f:

$$E[f(x)] = \sum_{i=1}^{M} \lambda_i f\left(x_i\right) = \left(\lambda_1 + \lambda_2\right)\left(\frac{\lambda_1 f\left(x_1\right) + \lambda_2 f\left(x_2\right)}{\lambda_1 + \lambda_2}\right) + \lambda_3 f\left(x_3\right) + \ldots + \lambda_M f\left(x_M\right)$$

$$\geq \left(\lambda_1 + \lambda_2\right) \cdot f\left(\frac{\lambda_1 x_1 + \lambda_2 x_2}{\lambda_1 + \lambda_2}\right) + \lambda_3 f\left(x_3\right) + \ldots + \lambda_M f\left(x_M\right)$$

$$\geq f\left(\left(\lambda_1 + \lambda_2\right) \cdot \left(\frac{\lambda_1 x_1 + \lambda_2 x_2}{\lambda_1 + \lambda_2}\right) + \lambda_3 x_3 + \ldots + \lambda_M x_M\right)$$

$$= f\left(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \ldots + \lambda_M x_M\right) = f(E[x])$$

Hence: E[f(x)] ≥ f(E[x])

Applications of Jensen's Inequality $E(X^2) \geq (E(X))^2$

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

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

Consider: Jensen's inequality states that for any convex function f

$$E_{avg} = \frac{1}{M} \sum_{i=1}^{M} E\left(\in_i(x)^2\right) \geq \frac{1}{M} \sum_{i=1}^{M} \left(E\left(\in_i(x)\right)\right)^2 = E\left[\left\{\frac{1}{M} \sum_{i=1}^{M} \in_i(x)\right\}^2\right] = E_{agg}$$

We can say that $E_{agg} \leq E_{avg}$ (proved)

3.

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

Also recall that the weight for the point i at step t+1 is given by:

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

$D_t(i)$:) is the normalized weight of point i in step t

$h_t(i)$ is the hypothesis (prediction) at step t for point i

α_t is the final "voting power" of hypothesis h_t

$y(i)$ is the true label for point i

Z_t is the normalization factor at step t (it ensures that the weights sum up to 1.0)

Note that at step 1, the points have equal weight

$$D_1 = \frac{1}{N}$$

N is the total number of data points.

At each of the steps, the total error of ht will be defined as $\in_t = \frac{1}{2} - \gamma_t$

Error for Adaboost can be measure with respect to weight D_t .

$$\in = \sum_{i=1:h(x_i) \neq y_i}^{M} D_t(i)$$

While $h_t(x_i)$ and y_i both in {1;-1}

$$D_{t+1}(i) = \frac{D_1(i)}{Z_1} \cdot e^{-\alpha_1 h_1(i) y(i)} . \frac{e^{-\alpha_2 h_2(i) y(i)}}{Z_2} \ldots \frac{e^{-\alpha_i h_i(i) y(i)}}{Z_i}$$

$$= \frac{1}{N} \frac{e^{\sum_{i=1}^{t} -\alpha_i h_i(i) y(i)}}{\prod_{i=1}^{t} Z_i} = \frac{1}{N} \frac{e^{-f_t(i) y(i)}}{\prod_{i=1}^{t} Z_i} with\ f_t(i) = \sum_{i=1}^{t} -\alpha_i h_i(i)$$

$$Hence: e^{-f_t(i) y(i)} = N\left(\prod_{i=1}^{t} Z_i\right) \sum_t D_{t+1}(i)$$

$$with \sum_t D_{t+1}(i) = 1$$

Now Total training error of H(x)

$$T_H = \frac{1}{N} \sum_{i:\, H(i) \neq y(i)} 1$$

$H(i) = sign(f(i))$ so

$$T_H = \frac{1}{N} \sum_{i:\, f(i)y(i) \le 0} 1 \le \sum_i e^{-f(i)y(i)} \; when \; e^{-Z} \ge 1 \; when \; Z \le 0$$

$$Hence: T_H \le \sum_i e^{-f(i)y(i)}$$

$$T_H \le \frac{1}{N}\left(N \prod_t Z_t\right) \sum_i D_{t+1}(i) = \prod_t Z_t$$

Now,

$$Z_t = \sum_i D_t(i) \cdot e^{-\alpha_t h_t(i) y(i)} = \sum_{i:\, h_t(i) = y(i)} D_t(i) \cdot e^{-\alpha_t} + \sum_{i:\, h_t(i) \neq y(i)} D_t(i) \cdot e^{\alpha_t}$$

$$h_t(i) = y(i) \; then \begin{cases} h_t(i) = 1 \\ y(i) = 1 \end{cases} or \begin{cases} h_t(i) = -1 \\ y(i) = -1 \end{cases}$$

$$Hence: h_t(i) \cdot y(i) = 1$$

$$h_t(i) \neq y(i) \; then \begin{cases} h_t(i) = 1 \\ y(i) = -1 \end{cases} or \begin{cases} h_t(i) = 1 \\ y(i) = -1 \end{cases}$$

$$Hence: h_t(i) \cdot y(i) = -1$$

Now consider: $\sum_{i:\, h_t(i) \neq y(i)} D_t(i) = \in_t ; \sum_{i:\, h_t(i) = y(i)} D_t(i) = 1 - \in_t$

$$Z_t = e^{-\alpha_t}\left(1 - \in_t\right) + e^{\alpha_t} \in_t$$

Prove that at the end of T steps, that mean minimize error T_H with α_t come to be $\frac{1}{2}\ln\left(\frac{1 - \in_t}{\in_t}\right)$

Hence: $Z_t = e^{-\alpha_t}\left(1 - \in_t\right) + e^{\alpha_t} \in_t = 2\sqrt{\in_t\left(1 - \in_t\right)}$ with $\in_t = \frac{1}{2} - \gamma_t$

$$Z = \sqrt{1 - 4\gamma_t}$$

$$1 + x \le e^x \; let \; x = -4\gamma_t^2$$

$$Then \; 1 - 4\gamma_t^2 \le e^{-4\gamma_t^2}$$

$$Hence: Z_t \le \sqrt{e^{-4\gamma_t^2}} = e^{-2\gamma_t^2}$$

Consider:

$$T_H \le \prod_t Z_t \le \prod_t e^{-2\gamma_t{}^2} \le e^{-2\sum_{t=1}^{M}\gamma_t{}^2}$$ (proved)