Lecture Fys-Stk3155/4155, November 23, 2023

Basics of boosting methods

Example: MSE

$$MSE = \frac{1}{n} \sum_i \left(y_i - f(x_i) \right)^2$$

Basic philosophy: improve $f(x_i)$ iteratively, starting with a simple method

$$f(x_i) = f_M(x_i) = \sum_{m=1}^{M} \beta_m b(x_i; \gamma_m)$$

$$\hat{\beta}, \hat{\gamma} = \underset{\beta, \gamma}{\arg\min}\ C(\beta, \gamma)$$

$m = 1$ $\quad f_1(x) = f_0(x) + \beta_1 b(x; \gamma_1)$

assume $f_0(x) = 0$

Example $b(x; \gamma) = 1 + \gamma x$

For every iteration $- m -$

$$\hat{\beta}_m, \hat{\gamma}_m = \underset{\beta, \gamma}{\arg\min}$$

$$\frac{1}{n} \sum_i \left[\left(y_i - f_{m-1}(x_i) \right) - \beta b(x_i; \gamma) \right]^2$$

in this example

$$\frac{1}{n} \sum_i \left(y_i - f_m(x_i) - \beta (1 + \gamma x_i) \right)^2$$

For every m

$$\frac{\partial C}{\partial \beta} = -\frac{2}{n} \sum_i (1+\gamma x_i)(y_i - f_{m-1}(x_i) - \beta(1+\gamma x_i)) = 0$$

$$\frac{\partial C}{\partial \gamma} = -\frac{2}{n} \sum_i \beta x_i (y_i - \beta(1+\gamma x_i)) = 0$$

with $\hat{\beta}_m$ and $\hat{\gamma}_m$

$$f_m(x) = f_{m-1}(x) + \hat{\beta}_m b(x; \hat{\gamma}_m)$$

Loop from $m=1$ to $M=m$

Algorithm

initialize $f_0(x;y)$ (often 0)

Define cost function and

M and $b(x,\gamma)$

initialize γ and β

for $m = 1 : M$

a) optimize and compute

$$(\hat{\beta}_m, \tilde{\gamma}_m) = \underset{\beta,\gamma}{\arg\min}\, C(\beta,\gamma)$$

b) set

$$f_m(x) = f_m(x;\gamma,\beta)$$
$$= f_{m-1}(x;\gamma_{m-1},\beta_{m-1}) + \beta_m b(x;\gamma_m)$$

end For

return $f_M(x)$

Boosting and classification

$$D = \{ (x_0, y_0), (x_1, y_1) \dots (x_{n-1}, y_{n-1}) \}$$

$$y_i = \{-1, +1\}$$

set of weak classifiers

$$\{ b_1, b_2 \dots b_M \}$$

$$b_j(x_i) \in \{-1, +1\}$$

$$C_m(x_i) = C_{m-1}(x_i) + \alpha_m b_m(x_i)$$

$$C_M(x) = \sum_{i=1}^{M} c_i(x)$$

How do we find α_m?

Define an error

$$E_{err} = \sum_{i=0}^{n-1} e^{-y_i C_m(x_i)}$$

$$y_i = \{-1, +1\} \wedge C_m(x_i) = \{-1, 1\}$$

correctly classified $C_m(x)\, y = +1$

wrongly — $\llcorner$ — $C_m(x)\, y = -1$

$$E_{err} = \sum_{i=0}^{n-1} \underbrace{e^{-y_i C_{m-1}(x_i)}}_{w_i^{(m)}} e^{-y_i \alpha_m b_m(x_i)}$$

$$= \sum_{i=0}^{n-1} w_i^{(m)} e^{-y_i \alpha_m b_m(x_i)}$$

split into correctly and wrongly classified

$$= \sum_{y_i = b_m(x_i)} w_i^{(m)} e^{-\alpha_m} + \sum_{y_i \neq b_m(x_i)} w_i^{(m)} e^{+\alpha_m}$$

starting value $w_i^{(0)} = 1$

$$\frac{d\, E_{err}}{d\, \alpha_m} = 0$$

$$= - \sum_{y_i = b_m(x_i)} w_i^{(m)} e^{-\alpha_m}$$

$$+ \sum_{y_i \neq b_m(x_i)} w_i^{(m)} e^{\alpha_m}$$

$$e^{-\alpha_m} \sum_{y_i = b_m(x_i)} w_i^{(m)} = e^{\alpha_m} \sum_{y_i \neq b_m} w_i^{(m)}$$

$$\alpha_m = \frac{1}{2} \log \left[\frac{\sum_{y_i = b_m} w_i^{(m)}}{\sum_{y_i \neq b_m} w_i^{(m)}} \right]$$

Define error rate

$$\varepsilon_m = \frac{\sum_{y_i \neq b_m} w_i^{(m)}}{\sum_{i=0}^{n-1} w_i^{(m)}}$$

$$\alpha_m = \frac{1}{2} \log \left[\frac{1 - \varepsilon_m}{\varepsilon_m} \right]$$

Algo for Adaboost

Data set $D = \{(x_0, y_0), (x_1, y_1) \cdots (x_{n-1}, y_{n-1})\}$

initialize $w_i^{(0)} = 1$

Define error function (cost function)

Define weak learner k_m

for $i = 1 : M$

optimize $\varepsilon_{err} \rightarrow \alpha_m$

$$\alpha_m = \frac{1}{2} \log \left[\frac{1 - \varepsilon_m}{\varepsilon_m} \right]$$

update $C_m(x) = C_{m-1}(x) + \alpha_m k_m(x)$

update weights

$$w_i^{(m+1)} = w_i^{(m)} e^{-y_i \alpha_m f_m(x)}$$

normalize weights

$$\sum_{i=0}^{n-1} w_i^{(m+1)} = 1$$

end for

Return $C_M(x)$