מבוא למערכות לומדות - גיליון קצר 5

מגיש - עומר שמחי, 316572593

20 ביוני 2021

חלק א' - רגרסייה

נזכר בבעיית הריבועים המינימאליים:

$$argmin_{w\in\mathbb{R}^d}\mathcal{L}\left(w\right)=argmin_{w\in\mathbb{R}^d}\left\Vert Xw-y\right\Vert _2^2=$$
$$=argmin_{w\in\mathbb{R}^d}\sum_{i=1}^{m}\left(w^Tx_i-y_i\right)^2,\quad y\in\mathbb{R}^m,X\in\mathbb{R}^{m\times d}$$

1. כעת נוכיח כי פונקציית ה-$loss$ היא פונקציה קמורה. לשם כך ראשית נרשום:

$$w=\begin{pmatrix}w_1\\ w_2\\ .\\ .\\ w_d\end{pmatrix},\ X=\begin{pmatrix}x_{11} & x_{12} & . & . & x_{1d}\\ x_{21} & x_{22} & . & . & x_{2d}\\ . & . & . & . & .\\ . & . & . & . & .\\ x_{m1} & x_{m2} & . & . & x_{md}\end{pmatrix}$$

(א) מצאו את הנגזרת ממעלה ראשונה $\frac{\partial}{\partial w_k}\mathcal{L}\left(w\right)$.

פתרון - נרשום:

$$\frac{\partial}{\partial w_k}\mathcal{L}\left(w\right)=\frac{\partial}{\partial w_k}\sum_{i=1}^{m}\left(w^Tx_i-y_i\right)^2=\frac{\partial}{\partial w_k}\sum_{i=1}^{m}\left(\begin{pmatrix}w_1 & w_2 & . & . & w_d\end{pmatrix}\begin{pmatrix}x_{i1}\\x_{i2}\\.\\.\\x_{id}\end{pmatrix}-y_i\right)^2$$

$$=\frac{\partial}{\partial w_k}\sum_{i=1}^{m}\left(\sum_{t=1}^{d}w_tx_{it}-y_i\right)^2=\sum_{i=1}^{m}\frac{\partial}{\partial w_k}\left(\sum_{t=1}^{d}w_tx_{it}-y_i\right)^2=$$

$$\underbrace{=}_{chain\,rule}\sum_{i=1}^{m}2\cdot\left(\sum_{t=1}^{d}w_tx_{it}-y_i\right)\cdot x_{ik}=\boxed{2\cdot\sum_{i=1}^{m}x_{ik}\left(\sum_{t=1}^{d}w_tx_{it}-y_i\right)}$$

(ב) מצאו את הנגזרת השנייה $\frac{\partial^2}{\partial w_j\partial w_k}\mathcal{L}\left(w\right)$.

פתרון - נשתמש בתוצאה מהסעיף הקודם ונקבל:

$$\frac{\partial^2}{\partial w_j\partial w_k}\mathcal{L}\left(w\right)=\frac{\partial}{\partial w_j}\left(2\cdot\sum_{i=1}^{m}x_{ik}\left(\sum_{t=1}^{d}w_tx_{it}-y_i\right)\right)=$$

$$=2\cdot\sum_{i=1}^{m}x_{ik}\frac{\partial}{\partial w_j}\left(\sum_{t=1}^{d}w_tx_{it}-y_i\right)=\boxed{2\cdot\sum_{i=1}^{m}x_{ij}\cdot x_{ik}}\ (*)$$

(ג) הסק כי מטריצת ההסיאן המתקבלת היא:

$$\nabla_w^2\mathcal{L}\left(w\right)=2X^TX$$

הוכחה - נרשום:

$$2X^TX=2\cdot\begin{pmatrix}x_{11} & x_{21} & . & . & x_{m1}\\x_{12} & x_{22} & . & . & x_{m2}\\. & . & . & . & .\\. & . & . & . & .\\x_{1d} & x_{2d} & . & . & x_{md}\end{pmatrix}\begin{pmatrix}x_{11} & x_{12} & . & . & x_{1d}\\x_{21} & x_{22} & . & . & x_{2d}\\. & . & . & . & .\\. & . & . & . & .\\x_{m1} & x_{m2} & . & . & x_{md}\end{pmatrix}$$

$$\Longrightarrow\left(2X^TX\right)_{jk}=2\cdot\sum_{i=1}^{m}x_{ij}\cdot x_{ik}$$

בדיוק כמו ב-$(*)$ ולכן נקבל כי אכן $\nabla_w^2\mathcal{L}\left(w\right)=2X^TX$.

(ד) פונקציה רב מימדית הינה קמורה אמ"מ מטריצת ההסיאן שלה חיובית חצי מוגדרת. נשים לב כי מטריצת ההסיאן הנ"ל הינה מטריצה גרהם (מטריצה מהצורה A^TA כאשר $A=\sqrt{2}X$) ולכן ככזו היא מטריצה חיובית חצי מוגדרת ולכן גם ההסיאן, כנדרש. מכאן נסיק כי הפונקציה $\mathcal{L}\left(w\right)$ אכן קמורה.

2. הוכיחו כי מתקיים:

$$\underbrace{argmax_{w\in\mathbb{R}^d}\prod_{i=1}^{m}\mathbb{P}\left(y_i,x_i;w\right)}_{Maximum-Likehood\,Estimator}=\underbrace{argmin_{w\in\mathbb{R}^d}\frac{1}{m}\cdot\sum_{i=1}^{m}\left|w^Tx_i-y_i\right|}_{Least\,absolute\,deviation}=argmin_{w\in\mathbb{R}^d}\mathcal{L}_{abs}\left(w\right)$$

כאשר:

$$\mathbb{P}\left(w_j\mid\mu,b\right)=\frac{1}{2b}e^{-\frac{|w_j-\mu|}{b}}$$
$$\mathbb{E}\left(w_j\right)=\mu,\ Var\left(w_j\right)=2b^2$$

הוכחה - נפתח את הנוסחה בצד שמאל לזאתי שבצד ימין לפי ההדרכה ובדומה השלבים שביצענו בהרצאה:

$$\boldsymbol{argmax_{w\in\mathbb{R}^d}\prod_{i=1}^{m}\mathbb{P}\left(y_i,x_i;w\right)}\underbrace{=}_{Chain\,rule}argmax_{w\in\mathbb{R}^d}\prod_{i=1}^{m}\mathbb{P}\left(y_i\mid x_i;w\right)\cdot\mathbb{P}\left(x_i\mid w\right)=$$
$$\underbrace{=}_{x_i\,and\,w\,independent}argmax_{w\in\mathbb{R}^d}\prod_{i=1}^{m}\mathbb{P}\left(y_i\mid x_i;w\right)\cdot\mathbb{P}\left(x_i\right)\underbrace{=}_{\prod_i\mathbb{P}(x_i)\,scalar}$$
$$=argmax_{w\in\mathbb{R}^d}\prod_{i=1}^{m}\mathbb{P}\left(y_i\mid x_i;w\right)\underbrace{=}_{log\,preserves\,argmax}$$
$$=argmax_{w\in\mathbb{R}^d}\log\left(\prod_{i=1}^{m}\mathbb{P}\left(y_i\mid x_i;w\right)\right)\underbrace{=}_{"\log\prod=\sum\log"}$$
$$=argmax_{w\in\mathbb{R}^d}\left(\sum_{i=1}^{m}\log\left(\mathbb{P}\left(y_i\mid x_i;w\right)\right)\right)=$$
$$\underbrace{=}_{\epsilon\sim Laplace(0,b)}argmax_{w\in\mathbb{R}^d}\left(\sum_{i=1}^{m}\log\left(\frac{1}{2b}e^{-\frac{\left|w^Tx_i-y_i\right|}{b}}\right)\right)=$$
$$=argmax_{w\in\mathbb{R}^d}\left(\sum_{i=1}^{m}\log\left(\frac{1}{2b}\right)-\frac{\left|w^Tx_i-y_i\right|}{b}\right)=$$
$$\underbrace{=}_{independent\,with\,w}argmax_{w\in\mathbb{R}^d}\left(-\frac{\left|w^Tx_i-y_i\right|}{b}\right)=-\frac{1}{b}argmax_{w\in\mathbb{R}^d}\left|w^Tx_i-y_i\right|=$$
$$\underbrace{=}_{b\,is\,a\,scalar}argmax_{w\in\mathbb{R}^d}\left|w^Tx_i-y_i\right|\underbrace{=}_{scalar}argmax_{w\in\mathbb{R}^d}\frac{1}{m}\cdot\left|w^Tx_i-y_i\right|=$$
$$:=\boldsymbol{argmin_{w\in\mathbb{R}^d}\mathcal{L}_{abs}\left(w\right)}\blacksquare$$

חלק ב' - $boosting$

3. הוכיחו כי מתקיים:

$$\sum_i D_i^{(t+1)} \cdot 1_{h_t(x_i)\neq y_i} = \frac{1}{2}$$

הוכחה - נפתח את אגף ימין תוך שימוש בשלבי ובהגדרות המשתנים באלגוריתם $AdaBoost$:

$$\sum_{\boldsymbol{i}} \boldsymbol{D_i^{(t+1)} \cdot 1_{h_t(x_i)\neq y_i}} = \sum_i D_i^{(t)} \cdot \frac{e^{-\alpha_t y_i h_t(x_i)}}{\sum_j D_j^{(t)} e^{-\alpha_t y_i h_t(x_j)}} \cdot 1_{h_t(x_i)\neq y_i} =$$

$$\underbrace{=}_{if\, h_t(x_i)\neq y_i\, then\, y_i h_t(x_i)=-1} \frac{1}{\sum_j D_j^{(t)} e^{-\alpha_t y_i h_t(x_j)}} \sum_{i,\, s.t\, y_i\neq h_t(x_i)} D_i^{(t)} \cdot e^{\alpha_t} =$$

$$= \frac{e^{\alpha_t} \cdot \sum_{i\, s.t\, y_i\neq h_t(x_i)} D_i^{(t)}}{\sum_{j,\, s.t\, y_i\neq h_t(x_i)} D_j^{(t)} e^{\alpha_t} + \sum_{j,\, s.t\, y_i = h_t(x_i)} D_j^{(t)} e^{-\alpha_t}} =$$

$$= \frac{e^{\alpha_t} \cdot \sum_i D_i^{(t)} \cdot 1_{h_t(x_i)\neq y_i}}{\sum_j D_j^{(t)} e^{\alpha_t} \cdot 1_{h_t(x_i)\neq y_i} + \sum_{j,\, s.t\, y_i = h_t(x_i)} D_j^{(t)} e^{-\alpha_t}} =$$

$$\underbrace{=}_{def.\, of\, \epsilon_t} \frac{e^{\alpha_t} \cdot \epsilon_t}{e^{\alpha_t} \cdot \epsilon_t + (1-\epsilon_t)\, e^{-\alpha_t}} = \frac{\epsilon_t}{\epsilon_t + \frac{1-\epsilon_t}{e^{2\alpha_t}}} =$$

$$\underbrace{=}_{def.\, of\, \alpha_t} \frac{\epsilon_t}{\epsilon_t + \frac{1-\epsilon_t}{e^{2\cdot\frac{1}{2}\cdot\log\left(\frac{1}{\epsilon t}-1\right)}}} = \frac{\epsilon_t}{\epsilon_t + \frac{1-\epsilon_t}{e^{\log\left(\frac{1}{\epsilon t}-1\right)}}} =$$

$$= \frac{\epsilon_t}{\epsilon_t + \frac{1-\epsilon_t}{\frac{1}{\epsilon t}-1}} = \frac{\epsilon_t}{\epsilon_t + \frac{1-\epsilon_t}{\frac{1-\epsilon_t}{\epsilon t}}} = \frac{\epsilon_t}{2\epsilon_t} = \boldsymbol{\frac{1}{2}} \blacksquare$$