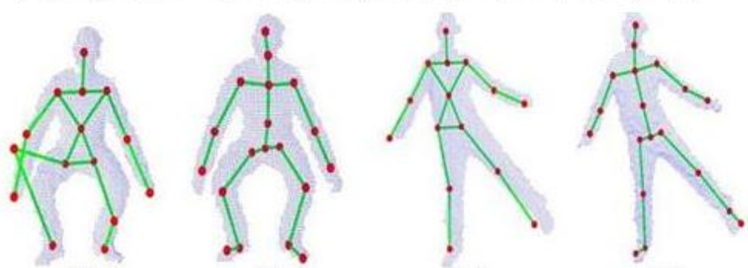
بسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

گروه سیستم‌های دیجیتال



آزمایشگاه یادگیری و بینایی ماشین

دستور کار آزمایش دوم: رگرسیون خطی

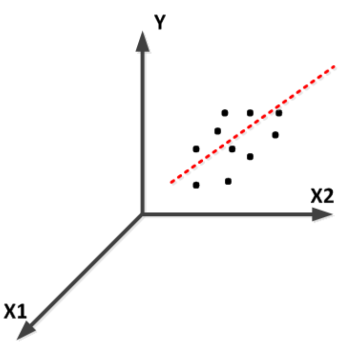
زمان لازم برای انجام آزمایش: حداکثر دو جلسه

Sajjad Hashembeiki

98107077

قسمت اول: رگرسیون خطی با تعداد ویژگی های محدود

منظور از رگرسیون خطی این است که با دانستن نقاطی نمونه از داده آماری، بتوان رفتار آن داده را در سایر نقاط و یا در آینده پیش‌بینی کرد. برای مثال، فرض کنید که کمیت (y) ما آلودگی هوا است. این آلودگی هوا به عواملی از جمله میزان تردد (x_1) و وسعت فضای سبز (x_2) وابسته است. در این صورت، اگر از تعدادی از نقاط شهر، نمونه‌ای از آلودگی هوا را بگیریم، می‌توانیم نمودار زیر را تشکیل دهیم.



همان‌طور که دیده می‌شود، تا وقتی که داده جمع‌آوری شده، کاملا نویز نباشد، می‌توان یک الگوی مشخصی از آن بدست آورد و یک خط با شیب غیر صفر به آن برازش کرد.

این خط برازش شده را در شکل، به صورت یک خط‌چین نمایش داده‌ایم.

اگر متغیرهای مساله را به صورت یک بردار $\boldsymbol{x}$ نمایش دهیم، خواهیم داشت:

$$\boldsymbol{x} = [x_1, x_2, \dots, x_p]^T$$

در این صورت، خروجی ما (که از آن نمونه گرفته‌ایم) به صورت زیر است:

$$y = f(\boldsymbol{x}) + \epsilon$$

که در آن، ε نویز است. هدف ما یافتن f(.) است که همان رگرسیون خطی داده است:

$$f(\boldsymbol{x}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

فرض کنید که N نمونه گرفته‌ایم یا داریم. که متغیرهای نمونه i ام به صورت زیر است:

$$\boldsymbol{x}^i = [x_1^i, x_2^i, \dots, x_p^i]^T$$

و خروجی آن (نمونه i ام) به صورت y^i است.

برای یافتن f ما باید ضرایب β را پیدا کنیم. باید β هایی را بیابیم که کمترین اختلاف بین y^iها و $f\left(\boldsymbol{x}^i\right)$ ها را ایجاد کند تا بهترین تخمین ممکن را داشته باشیم. ما برای نمایش تفاضل این دو مقدار، از تفاضل نُرم ۲ استفاده می‌کنیم:

$$RSS = \sum_{i=1}^{N} \left(y^i - f\left(\boldsymbol{x}^i\right)\right)^2$$

ما باید عبارت بالا را حداقل کنیم. می‌توان نشان داد که عبارت RSS بالا به بیان ماتریسی، به صورت زیر می‌شود:

$$RSS = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$$

که در آن، ماتریس $\mathbf{X}$ به صورت زیر است:

$$\mathbf{X} = \begin{bmatrix} 1 & {\boldsymbol{x}^1}^T \\ & \vdots \\ 1 & {\boldsymbol{x}^N}^T \end{bmatrix}_{N\times(P+1)}$$

و $\boldsymbol{y}$ بردار ستونی است که خروجی نمونه‌ها را نشان می‌دهد.

(امتیازی): رابطه ماتریسی RSS را اثبات نمایید.

اگر از رابطه RSS بالا نسبت به $\boldsymbol{\beta}$ مشتق بگیریم و مساوی صفر قرار دهیم (مساله بهینه‌سازی)، β به صورت زیر بدست می‌آید:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$$

از رابطه بالا برای تخمین $\boldsymbol{y}$ (خروجی نمونه های آموزشی) به صورت زیر استفاده می‌شود:

$$\hat{\boldsymbol{y}} = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$$

Bonus

$$RSS = \sum_{i=1}^{N} \left(y^i - f(\underline{x}^i) \right)^2$$

$$f(\underline{x}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1}, \quad \underline{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{bmatrix}_{(p+1)\times 1}, \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$$

$$\Rightarrow f(\underline{x}) = \underline{x}^T \underline{\beta}$$

$$RSS = \sum_{i=1}^{N} \left(y^i - \underline{x}^{i\,T} \underline{\beta} \right)^2$$

$$= \begin{bmatrix} y^1 - \underline{x}^{1T}\underline{\beta} & y^2 - \underline{x}^{2T}\underline{\beta} & \cdots & y^N - \underline{x}^{NT}\underline{\beta} \end{bmatrix}_{1\times N} \begin{bmatrix} y^1 - \underline{x}^{1T}\underline{\beta} \\ y^2 - \underline{x}^{2T}\underline{\beta} \\ \vdots \\ y^N - \underline{x}^{NT}\underline{\beta} \end{bmatrix}_{N\times 1} = Z^T Z$$

$$\underline{Z} = \begin{bmatrix} y^1 - \underline{x}^{1T}\underline{\beta} \\ y^2 - \underline{x}^{2T}\underline{\beta} \\ \vdots \\ y^n - \underline{x}^{nT}\underline{\beta} \end{bmatrix}_{n\times 1} = \underline{y} - \begin{bmatrix} \underline{x}^{1T}\underline{\beta} \\ \underline{x}^{2T}\underline{\beta} \\ \vdots \\ \underline{x}^{nT}\underline{\beta} \end{bmatrix}_{n\times 1}$$

$$= \underline{y} - \underset{n\times(p+1)}{\begin{bmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_p^1 \\ 1 & x_1^2 & x_2^2 & \cdots & x_p^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_1^n & x_2^n & \cdots & x_p^n \end{bmatrix}} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1}$$

$$= \underline{y} - \underline{X}\,\underline{\beta}$$

$$\Rightarrow RSS = Z^T Z \xrightarrow{Z = \underline{y} - \underline{X}\underline{\beta}} \boxed{RSS = (\underline{y} - \underline{X}\,\underline{\beta})^T(\underline{y} - \underline{X}\,\underline{\beta})}$$

پیش گزارش

۱- در یک مسألهٔ رگرسیون خطی، RSS را به صورت زیر تعریف می شود:

$$RSS = \sum_{i=1}^{n}(y^i - \beta_0 - \beta_1 x^i)^2$$

β_0^* و β_1^* جواب‌های این مسأله هستند. از بین معادله‌های زیر، آن‌هایی که صحیح هستند را مشخص کنید.

(راهنمایی: از معادلهٔ بالا نسبت به β_0 و β_1 مشتق بگیرید.)

False
$$\left\{\begin{array}{l} \sum_{i=1}^{n}\left(y^i - {\beta_0}^* - {\beta_1}^* x^i\right)y^i = 0 \\ \sum_{i=1}^{n}\left(y^i - {\beta_0}^* - {\beta_1}^* x^i\right)(y^i - \bar{y}) = 0 \end{array}\right.$$

True
$$\left\{\begin{array}{l} \sum_{i=1}^{n}\left(y^i - {\beta_0}^* - {\beta_1}^* x^i\right)(x^i - \bar{x}) = 0 \\ \sum_{i=1}^{n}\left(y^i - {\beta_0}^* - {\beta_1}^* x^i\right)({\beta_0}^* + {\beta_1}^* x^i) = 0 \end{array}\right.$$

۲- در دستور کار با مفهوم overfitting آشنا خواهید شد. در مورد underfitting تحقیق نمایید و کمی درباره این مفهوم توضیح دهید.

1) $$RSS = \sum_{i=1}^{N} \left(y^i - \beta_0 - \beta_1 x^i\right)^2$$

$$\rightarrow \frac{\partial RSS}{\partial \beta_0} = \sum_{i=1}^{N} -1 \times 2\left(y^i - \beta_0 - \beta_1 x^i\right)$$

$$\Rightarrow \underbrace{\sum_{i=1}^{N} \left(y^i - \beta_0^* - \beta_1^* x^i\right)}_{=A} = 0 \qquad \bigstar$$

$$\rightarrow \frac{\partial RSS}{\partial \beta_1} = \sum_{i=1}^{N} -x^i \times 2\left(y^i - \beta_0 - \beta_1 x^i\right)$$

$$\Rightarrow \underbrace{\sum_{i=1}^{N} x^i\left(y^i - \beta_0^* - \beta_1^* x^i\right)}_{=B} = 0 \qquad \bigstar\bigstar$$

The third equ $\longrightarrow \Rightarrow B - \bar{x} A \overset{B=0,\, A=0}{=\!=} 0$ ✓

The last equ $\longrightarrow \Rightarrow A\beta_0^* + \beta_1^* B \underset{B=0}{\overset{A=0}{=\!=}} 0$ ✓

2: Underfitting: model can't capture (learn) the relationship between inputs & outputs properly and error rate on training set & test set is high. Basically the model isn't complex enough. In this situation the model has High Bias and low variance.