



به نام خدا
دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر



درس یادگیری ماشین تمرین سوم

نام و نام خانوادگی	امیرحسین پورداود
شماره دانشجویی	۸۱۰۱۰۱۱۲۰
تاریخ ارسال گزارش	۱۴۰۲.۱۰.۰۳

فهرست

پاسخ ۱ - پرسش	۳
۱-۱. قسمت ۱	۳
۲-۱. قسمت ۲	۳
۳-۱. قسمت ۳	۴
۴-۱. قسمت ۴	۵
پاسخ ۲ - پرسش	۶
۱-۲. قسمت ۱	۶
۲-۲. قسمت ۲	۷
پاسخ ۳ - پرسش	۸
پاسخ ۴ - پرسش	۹
۱-۴. قسمت ۱	۹
۲-۴. قسمت ۲	۹
پاسخ ۵ - شبیه سازی	۱۰
۱-۵. قسمت ۱	۱۰
۲-۵. قسمت ۲	۱۰
۳-۵. قسمت ۳	۱۱
۴-۵. قسمت ۴	۱۱
۵-۵. قسمت ۵	۱۲
۷-۵. قسمت ۷	۱۲
۸-۵. قسمت ۸	۱۲
۹-۵. قسمت ۹	۱۲
۱۰-۵. قسمت ۱۰	۱۲

۱۱-۵	قسمت ۱۱	۱۳
۱۲-۵	قسمت ۱۲	۱۳
۱۳-۵	قسمت ۱۳	۱۳
پاسخ ۶- شبیه سازی		۱۴
۱-۶	قسمت ۱	۱۴
۲-۶	قسمت ۲	۱۴
۳-۶	قسمت ۳	۱۵
۴-۶	قسمت ۴	۱۵
۵-۶	قسمت ۵	۱۵
۶-۶	قسمت ۶	۱۶
۷-۶	قسمت ۷	۱۶
۸-۶	قسمت ۸	۱۶
۹-۶	قسمت ۹	۱۶
۱۰-۶	قسمت ۱۰	۱۷
پاسخ ۷ - شبیه سازی		۱۷
۱-۷	قسمت ۱	۱۷
۲-۷	قسمت ۲	۱۸
۲-۷	grid search	۱۸
۳-۷	random search	۱۹
پاسخ ۸ - شبیه سازی		۲۱

پاسخ ۱ - پرسش

۱-۱. قسمت ۱

۱-۱)

$$y = \beta_0 + \varepsilon_i$$

$$p(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) \Rightarrow p(y_i | x_i; \beta_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_0)^2}{2\sigma^2}\right)$$

$$L(\beta_0) = \prod_{i=1}^n p(y_i | x_i; \beta_0)$$

$$\max_{\beta_0} L(\beta_0) \equiv \max_{\beta_0} \log L(\beta_0) \equiv \min_{\beta_0} \sum_{i=1}^n (y_i - \beta_0)^2$$

$$\Rightarrow \frac{\partial}{\partial \beta_0} \sum_{i=1}^n (y_i - \beta_0)^2 = 0 \Rightarrow \sum_{i=1}^n -2(y_i - \beta_0) = 0$$

$$\Rightarrow \beta_0 = \frac{\sum_{i=1}^n y_i}{n}$$

$$\beta_0 = \frac{561}{10} = 56,1 \quad \checkmark$$

$$\Rightarrow \boxed{y = 56,1 + \varepsilon_i}$$

۲-۱. قسمت ۲

۱-۲)

$$y = \beta_1 x_i + \varepsilon_i$$

$$p(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) \Rightarrow p(y_i | x_i; \beta_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_1 x_i)^2}{2\sigma^2}\right)$$

$$L(\beta_1) = \prod_{i=1}^n p(y_i | x_i; \beta_1)$$

$$\max_{\beta_1} L(\beta_1) \equiv \max_{\beta_1} \log L(\beta_1) \equiv \min_{\beta_1} \sum_{i=1}^n (y_i - \beta_1 x_i)^2$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{10} (y_i - \beta_1 x_i)^2 = 0 \Rightarrow \sum_{i=1}^{10} x_i (y_i - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{10} x_i y_i = \beta_1 \sum_{i=1}^{10} x_i^2$$

$$\Rightarrow \beta_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\beta_1 = \frac{12521}{4173} \approx \frac{3}{1} \quad \checkmark$$

$$\rightarrow \boxed{y = 3x + \varepsilon_i}$$

۳-۱. قسمت ۳

۱-۳)

$$\hat{y} = 25 - 0.5x$$

$$x = 6 \rightarrow \hat{y} = 22$$

در صورت در نظر گرفتن خطا، مشاهده جدید یک توزیع نرمال با میانگین ۲۲ خواهد بود، پس لزوماً مقدار ۲۲ را مشاهده نخواهیم کرد.

۱-۴)

$$n=16 \rightarrow \sum_{i=1}^{16} (y_i - \hat{y}_i)^2 = 7$$

$$p(y_i | \hat{y}_i; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right) \Rightarrow L(\sigma) = \prod_{i=1}^n p(y_i | \hat{y}_i; \sigma)$$

$$\rightarrow \max_{\sigma} L(\sigma) \equiv \max_{\sigma} \log L(\sigma) \equiv \max_{\sigma} -n \log \sigma - \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{2\sigma^2}$$

۱-۴)

$$n=16 \rightarrow \sum_{i=1}^{16} (y_i - \hat{y}_i)^2 = 7$$

$$p(y_i | \hat{y}_i; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right) \Rightarrow L(\sigma) = \prod_{i=1}^n p(y_i | \hat{y}_i; \sigma)$$

$$\rightarrow \max_{\sigma} L(\sigma) \equiv \max_{\sigma} \log L(\sigma) \equiv \max_{\sigma} -n \log \sigma - \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{2\sigma^2}$$

$$\Rightarrow \min_{\sigma} n \log \sigma + \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{SSE}$$

$$\Rightarrow \frac{\partial}{\partial \sigma} \left(n \log \sigma + \frac{SSE}{2\sigma^2} \right) = 0 \Rightarrow \frac{n}{\sigma} - \frac{SSE}{\sigma^3} = 0$$

$$\Rightarrow n\sigma^2 - SSE = 0$$

$$\Rightarrow \sigma^2 = \frac{SSE}{n}$$

$$\begin{cases} SSE = 7 \\ n = 16 \end{cases} \Rightarrow \sigma^2 = \frac{7}{16} \approx \underline{\underline{0.44}} \quad \checkmark$$

پاسخ ۲ - پرسش

۲-۱. قسمت ۱



در شکل بالا، تابع sigmoid برای وزن های مختلف رسم شده است.

در logistic regression زمانی بیش برازش رخ میدهد که مدل داده های آموزشی را بیش از اندازه یاد بگیرد و در واقع نویز و نوسانات موجود در داده های آموزشی را هم یاد بگیرد. در وزن های بزرگ میزان گرادیان بسیار افزایش میکند. همچنین وزن های بزرگ مدل را نسبت به داده های پرت حساس تر میکند و نویز و منجر به یادگیری نویز و نوسانات نیز میشود.

وزن های بزرگ منجر به مدلی با واریانس بالا میشود که خود نشان از بیش برازش است.

وزن های بزرگ، میتوانند به مرز های تصمیم گیری پیچیده تری منجر بشوند و ممکن است مدل الگو هایی را یاد بگیرد که فقط منجر به داده های آموزشی شود و به داده های جدید تعمیم پیدا نکند.

Q2/

2-2)

$$\max_{\omega_0, \dots, \omega_d} \prod_{i=1}^n P(Y_i | X_i, \omega_0, \dots, \omega_d) P(\omega_0, \dots, \omega_d)$$

$$P(\omega_0, \dots, \omega_d) \sim \mathcal{N}(0, I) = \frac{1}{(2\pi)^{\frac{d+1}{2}}} \exp\left(-\frac{\omega_0^2 + \dots + \omega_d^2}{2}\right) = \prod_{i=0}^d \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\omega_i^2}{2}\right)$$

$$L(\omega) = \log \prod_{i=1}^n P(Y_i | X_i, \omega_0, \dots, \omega_d) P(\omega_0, \dots, \omega_d)$$

$$= \log P(\omega_0, \dots, \omega_d) + \sum_{i=1}^n \log P(Y_i | X_i, \omega_0, \dots, \omega_d)$$

$$= \sum_{j=0}^d \left(\log \frac{1}{\sqrt{2\pi}} - \frac{1}{2} \omega_j^2 \right) + \sum_{i=1}^n \log P(Y_i | X_i, \omega_0, \dots, \omega_d) ; P(Y_i | X_i, \omega) = \frac{1}{1 + e^{-\omega^T x_i}}$$

$$\omega_j^{k+1} \leftarrow \omega_j^k + \eta \frac{\partial L(\omega)}{\partial \omega_j}$$

$$\nabla_{\omega} L(\omega) = \underline{\omega} - \sum_{i=1}^n x_i (y_i - \text{sig}(\underline{\omega}^T x_i))$$

$$\omega^{k+1} = \omega^k - \eta \left(\sum_{i=1}^n -x_i (y_i - \text{sig}(\omega^k{}^T x_i)) \right) + \omega^k$$

$$\Rightarrow \omega^{k+1} = \omega^k (1 - \eta) + \eta \left[\sum_{i=1}^n x_i (y_i - \text{sig}(\omega^k{}^T x_i)) \right] \quad \checkmark$$

پاسخ ۳ - پرسش

$$x^1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad y_1 = -1$$

$$x^3 = \begin{pmatrix} -1 \\ -2 \end{pmatrix} \quad y_3 = 1$$

$$x^2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad y_2 = -1$$

$$SVM \rightarrow y = w_1 x_1 + w_2 x_2 + b = \underline{w}^T \underline{x} + b$$

$$\text{Min } \frac{1}{2} \|\underline{w}\|^2$$

$$\text{s.t. } 1 - y_i (\underline{w}^T \underline{x}_i + b) \leq 0$$

Dual problem:

$$\text{Max } L = \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \underline{x}_i^T \underline{x}_j$$

$$\text{s.t. } \lambda_i \geq 0, \quad \sum_{i=1}^n \lambda_i y_i = 0$$

$$L = \lambda_1 + \lambda_2 + \lambda_3 - \frac{1}{2} \sum_{i=1}^3 \lambda_i y_i [\lambda_1 y_1 \underline{x}_1^T \underline{x}_i + \lambda_2 y_2 \underline{x}_2^T \underline{x}_i + \lambda_3 y_3 \underline{x}_3^T \underline{x}_i]$$

$$= \lambda_1 + \lambda_2 + \lambda_3 - \frac{1}{2} [(\lambda_1 y_1 \|\underline{x}_1\|)^2 + (\lambda_2 y_2 \|\underline{x}_2\|)^2 + (\lambda_3 y_3 \|\underline{x}_3\|)^2$$

$$+ 2\lambda_1 \lambda_2 y_1 y_2 \underline{x}_2^T \underline{x}_1 + 2\lambda_1 \lambda_3 y_1 y_3 \underline{x}_3^T \underline{x}_1 + 2\lambda_2 \lambda_3 y_2 y_3 \underline{x}_3^T \underline{x}_2]$$

$$\underline{x}_1^T \underline{x}_3 = -5 \quad \underline{x}_1^T \underline{x}_2 = 3 \quad \underline{x}_2^T \underline{x}_3 = -3 \quad \|\underline{x}_1\| = \|\underline{x}_2\| = \|\underline{x}_3\| = \sqrt{5}$$

$$y_1 y_2 = 1 \quad y_1 y_3 = -1 \quad y_2 y_3 = -1$$

$$\Rightarrow \text{Min } n \log \sigma + \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{SSE}$$

$$\Rightarrow \frac{\partial}{\partial \sigma} (n \log \sigma + \frac{SSE}{2\sigma^2}) = 0 \Rightarrow \frac{n}{\sigma} - \frac{SSE}{\sigma^3} = 0$$

$$\Rightarrow n\sigma^2 - SSE = 0$$

$$\Rightarrow \sigma^2 = \frac{SSE}{n}$$

$$\begin{cases} SSE = 7 \\ n = 16 \end{cases} \Rightarrow \sigma^2 = \frac{7}{16} \approx \underline{\underline{0.44}} \quad \checkmark$$

پاسخ ۴ - پرسش

۴-۱. قسمت ۱

4-1)

زبان که داده‌های ما قابل جداسازی نباشند، می‌توانیم با انتقال داده‌ها به یک فضای بالاتر تحت یک تابع که به آن Kernel گوئیم، آن‌ها را به گونه‌ای که قابل جداسازی با SVM باشند تبدیل کنیم. در SVM نیز که یک جداساز خطی است، نیاز است که داده‌ها مجدداً پذیر باشند و این امر با انتقال به فضای ویژگی بالاتر ایجاد می‌شود.

۴-۲. قسمت ۲

4-2)

$$K(x, x') = K_1(x, x') K_2(x, x')$$

$$K(x, y) = \sum_i \phi_i(x) \phi_i(y)$$

$$K_1(x, y) K_2(x, y) = \left(\sum_i \phi_i'(x) \phi_i'(y) \right) \left(\sum_j \phi_j^2(x) \phi_j^2(y) \right)$$

$$= \sum_{i,j} \underbrace{\phi_i'(x) \phi_j^2(x)}_{\phi_k'(x)} \underbrace{\phi_i'(y) \phi_j^2(y)}_{\phi_k'(y)} = \sum_k \phi_k'(x) \phi_k'(y)$$

$$= K'(x, y) \quad \square$$

پاسخ ۵ - شبیه سازی

۵-۱. قسمت ۱

مشخصات آماری این دیتاست، بصورت زیر میباشد:

	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
count	500.000000	500.000000	500.000000	500.000000	500.000000
mean	33.053194	12.052488	37.060445	3.533462	499.314038
std	0.992563	0.994216	1.010489	0.999278	79.314782
min	29.532429	8.508152	33.913847	0.269901	256.670582
25%	32.341822	11.388153	36.349257	2.930450	445.038277
50%	33.082008	11.983231	37.069367	3.533975	498.887875
75%	33.711985	12.753850	37.716432	4.126502	549.313828
max	36.139662	15.126994	40.005182	6.922689	765.518462

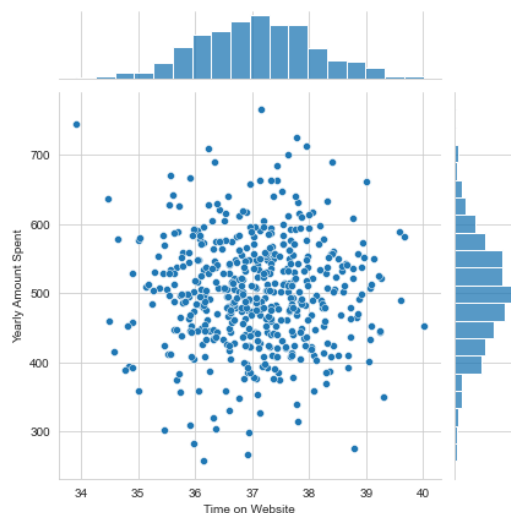
مشخصات عددی این دیتاست نظیر اسم کلاس ها و غیره نیز به این صورت میباشد:

```
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Email                  500 non-null    object  
1   Address                 500 non-null    object  
2   Avatar                  500 non-null    object  
3   Avg. Session Length    500 non-null    float64  
4   Time on App             500 non-null    float64  
5   Time on Website         500 non-null    float64  
6   Length of Membership    500 non-null    float64  
7   Yearly Amount Spent     500 non-null    float64  
dtypes: float64(5), object(3)
```

در این دیتاست ما ۸ کلاس داریم که هر کلاس، ۵۰۰ ردیف داده دارد.

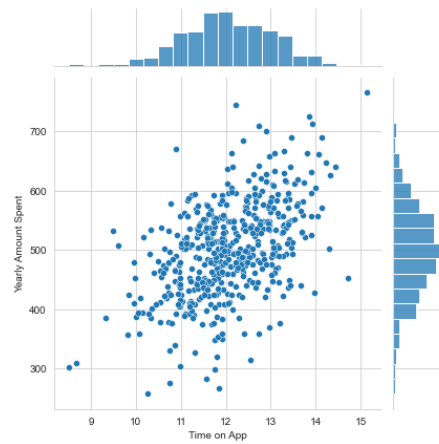
۵-۲. قسمت ۲

خروجی این قسمت، به صورت زیر میباشد:



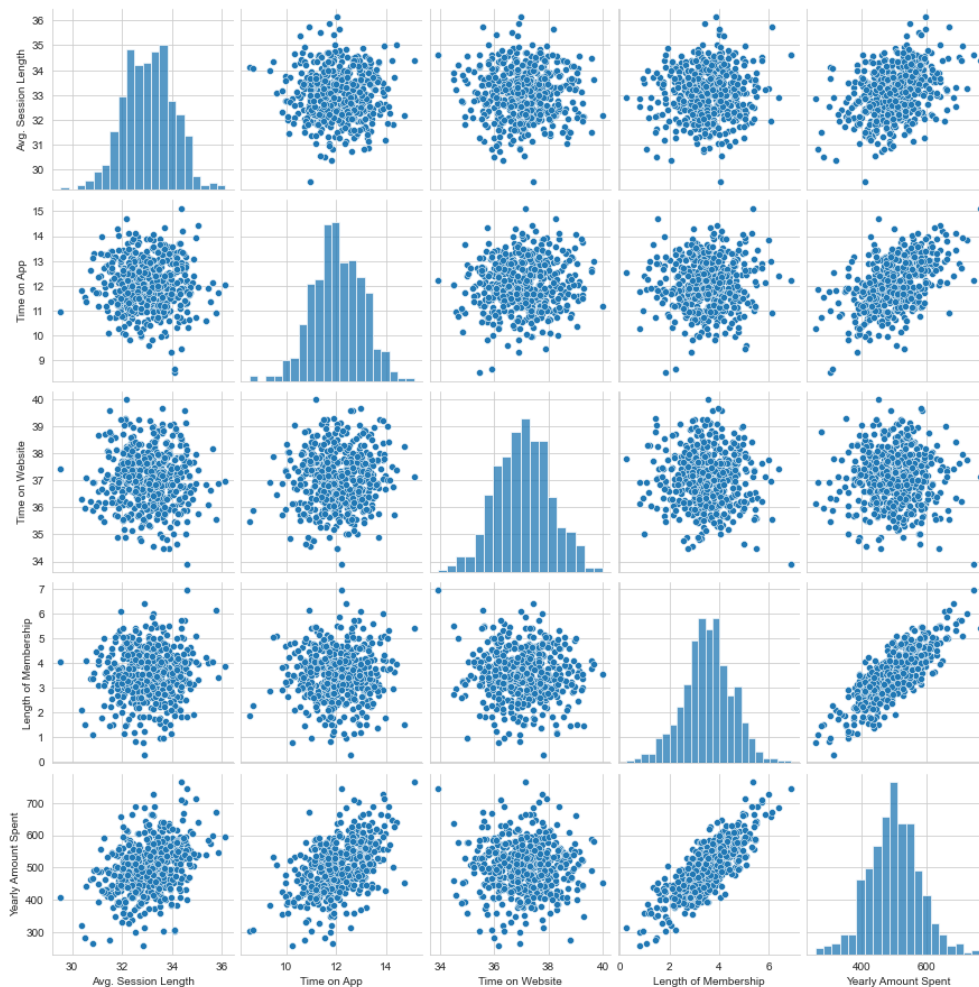
۳-۵. قسمت ۳

خروجی این قسمت، به صورت زیر میباشد:



۴-۵. قسمت ۴

بر اساس نمودار زیر میتوان گفت که میزان زمان عضویت در این شرکت ، مهمترین و خطی ترین عامل خرید میباشد .



۵-۵. قسمت ۵

در قسمت پیش پردازش، ابتدا داده هایی که به آن ها نیاز نداریم مانند آدرس، ایمیل و آواتار را از دیتاست حذف می‌نماییم و بر روی داده های باقیمانده نیز یک نرمالیزیشن انجام می‌دهیم تا در رنج مناسب ۰ و ۱ برای آموزش مدل قرار گیرند.

۵-۷. قسمت ۷

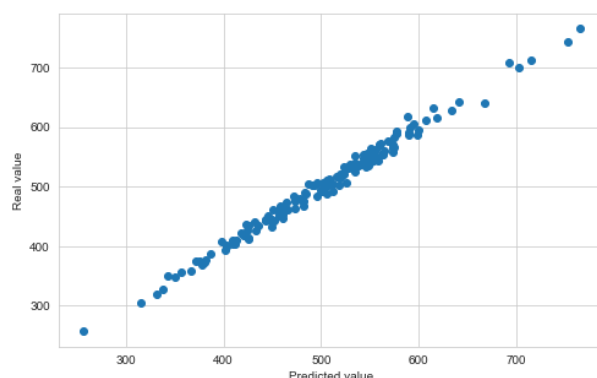
با استفاده از `linear_model.LinearRegression` یک مدل خطی روی داده ها آموزش می‌دهیم.

۵-۸. قسمت ۸

پس از آموزش مدل، بر روی داده های تست، پیش بینی انجام می‌دهیم که نتایج آن در قسمت های بعد آورده خواهد شد.

۵-۹. قسمت ۹

بر روی داده های تست و واقعی یک اسکتر پلات رسم می‌کنیم که خروجی آن بصورت زیر است:



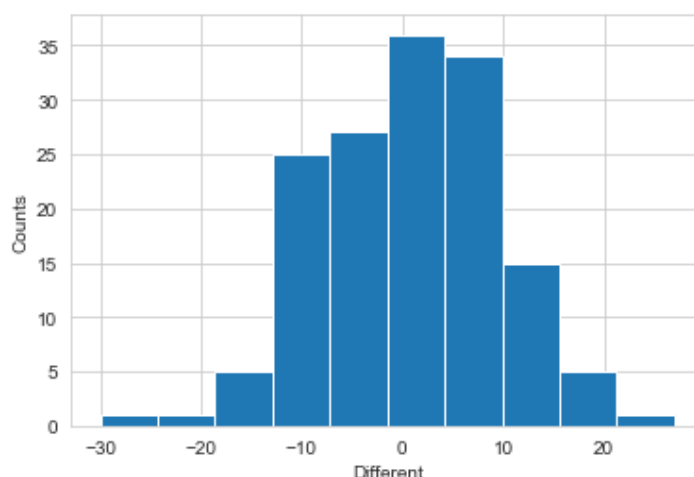
۵-۱۰. قسمت ۱۰

مقادیر خطاهای ذکر شده برای این مثال بصورت زیر خواهند بود:

```
Mean Square Error 79.81305165097454
Root Mean Square Error 8.933815066978639
Mean Absolute Error : 7.228148653430835
```

۵-۱۱. قسمت ۱۱

همانطور که در شکل زیر مشخص است به دلیل اینکه ما رگرسیون را بصورت میانگین خطا ایجاد شده برحسب MSE محاسبه میکنیم ، بنابراین خروجی بصورت ماینگین ۰ و بصورت اختلاف گوسی از داده های اصلی میتوانیم مشاهده کنیم.



۵-۱۲. قسمت ۱۲

با مشاهده جدول Coef. میتوان گفت که در صورت عدم نرمالایز داده ها، ضرایب مقادیر بسیار زیادی را اختیار میکردند و از طرفی نیز به مدل به داده های با مقادیر بالاتر بیشتر میل میکرد و همچنین امکان مقایسه برای ارزش ویژگی ها نیز وجود نداشت.

```
1 regr.coef_  
✓ 0.0s  
array([171.66616628, 255.42216789, 1.15982222, 407.67685316])
```

۵-۱۳. قسمت ۱۳

همانطور که مشخص است، ۴ امین فیچر یا همان زمان عضویت، بیشترین اثر در بالاترین خرید سالانه در این سایت داشته است. پس بهترین کار این است که شرکت، مشتری های خود را نگه دارد و اعتماد آن ها را جلب کند تا در عضویت بمانند و خارج نشوند.

در قسمت بعد نیز تاثیر نرم افزار در خرید بیشتر بوده است، یعنی با جذاب کردن اپ میتوان خریدار را ترغیب کرد خرید بیشتری انجام دهد.

پاسخ ۶- شبیه سازی

۶-۱. قسمت ۱

مشخصات آماری این دیتاست، بصورت زیر می باشد:

	Daily Time Spent on Site	Age	Area Income	Daily Internet Usage	Male	Clicked on Ad
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	65.000200	36.009000	55000.000080	180.000100	0.481000	0.500000
std	15.853615	8.785562	13414.634022	43.902339	0.499889	0.500250
min	32.600000	19.000000	13996.500000	104.780000	0.000000	0.000000
25%	51.360000	29.000000	47031.802500	138.830000	0.000000	0.000000
50%	68.215000	35.000000	57012.300000	183.130000	0.000000	0.500000
75%	78.547500	42.000000	65470.635000	218.792500	1.000000	1.000000
max	91.430000	61.000000	79484.800000	269.960000	1.000000	1.000000

مشخصات عددی این دیتاست نظیر اسم کلاس ها و غیره نیز به این صورت می باشد:

```

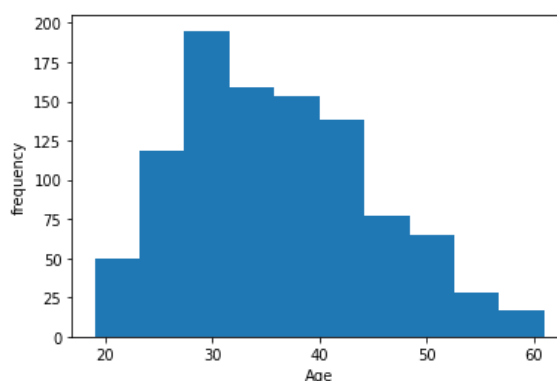
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype  
---  -
0   Daily Time Spent on Site              1000 non-null   float64
1   Age                                    1000 non-null   int64   
2   Area Income                           1000 non-null   float64
3   Daily Internet Usage                  1000 non-null   float64
4   Ad Topic Line                         1000 non-null   object  
5   City                                   1000 non-null   object  
6   Male                                   1000 non-null   int64   
7   Country                               1000 non-null   object  
8   Timestamp                             1000 non-null   object  
9   Clicked on Ad                         1000 non-null   int64   
dtypes: float64(3), int64(3), object(4)

```

در این دیتاست ما ۱۰ کلاس داریم که هر کلاس، ۱۰۰۰ ردیف داده دارد.

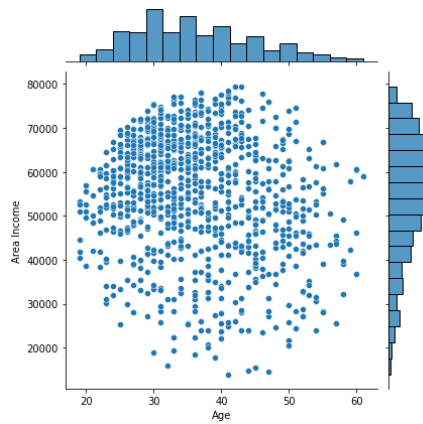
۶-۲. قسمت ۲

نمودار توزیع آماری برحسب سن بصورت زیر می باشد:



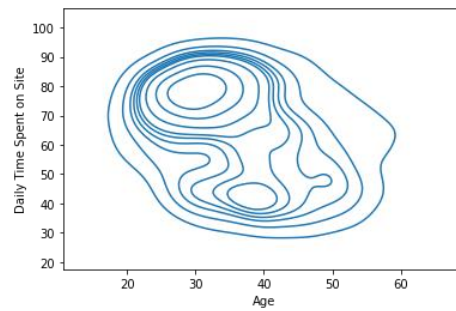
۳-۶. قسمت ۳

جوینت پلات درآمد بر حسب سن بصورت زیر میباشد:

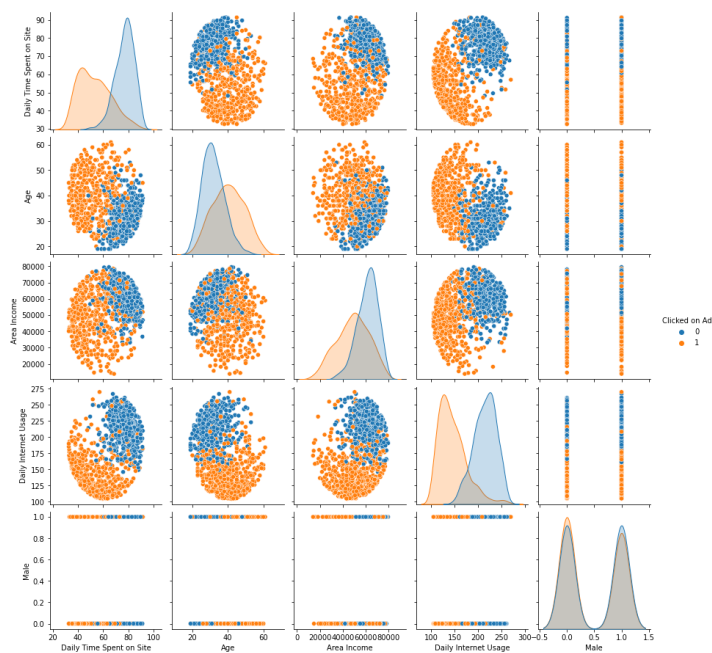


۴-۶. قسمت ۴

KDE پلات میزان زمان گذاشته شده روی سایت تبلیغ بر حسب سن بصورت زیر میباشد:



۵-۶. قسمت ۵



۶-۶. قسمت ۶

همانطور که از خروجی مشاهده می‌شود، کسانی که حجم اینترنتی کمتری مصرف میکنند، کلیک بیشتری انجام میدهند شاید به این علت باشد که وقتی وارد اینترنت می‌شوند با عدم آگاهی کافی و نداشتن تجربه کلیک میکنند

۶-۷. قسمت ۷

در این قسمت پیش پردازش بر روی داده‌ها انجام میدهم و آن‌ها را برای آموزش و تست مدل آماده میکنیم. ابتدا خالی بودن و Nan بودن داده‌ها را چک کرده و سپس ویژگی‌های غیر نیاز مانند 'Timestamp'، 'Country'، 'Male'، 'City' و 'Ad topic line' را حذف کرده و در نهایت داده‌ها را بین ۰ و ۱ نرمالایز میکنیم.

۶-۸. قسمت ۸

با توجه به پیاده سازی تابع لجستیک در مسئله، داده‌های ورودی مسئله را آموزش میدهم و ضرایب را بدست می‌آوریم.

```
[[ -4.27596562  
 [ 9.16332741  
 [ 1.32409984  
 [-5.07791832]]
```

۶-۹. قسمت ۹

با بدست آوردن مدل، بر روی داده‌های تست نیز پیشبینی انجام میدهم و دقت‌های زیر را بدست آورده ایم:

```
the Confusion matrix is :  
[[138  7]  
 [ 20 135]]  
Accuracy score is: 0.91
```

۶-۱۰. قسمت ۱۰

```
{'0': {'precision': 0.8734177215189873,
'recall': 0.9517241379310345,
'f1-score': 0.9108910891089109,
'support': 145},
'1': {'precision': 0.9507042253521126,
'recall': 0.8709677419354839,
'f1-score': 0.9090909090909091,
'support': 155},
'accuracy': 0.91,
'macro avg': {'precision': 0.91206097343555,
'recall': 0.9113459399332592,
'f1-score': 0.90999099909991,
'support': 300},
'weighted avg': {'precision': 0.9133490818327689,
'recall': 0.91,
'f1-score': 0.9099609960996099,
'support': 300}}
```

پاسخ ۷ - شیه سازی

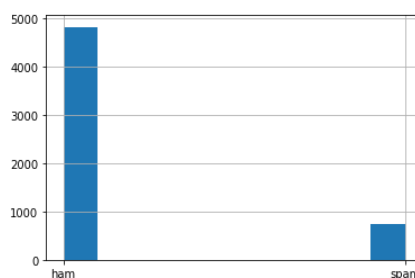
۷-۱. قسمت ۱

در این قسمت دیتاست را لود میکنیم و خلاصه ای از وضعیت آن ها و تعداد برچسب و کلاس ها را مشاهده میکنیم.

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
count	5572	5572	50	12	6
unique	2	5169	43	10	5
top	ham	Sorry, I'll call later	bt not his girlfrnd... G o o d n i g h t . . @"	MK17 92H. 450Ppw 16"	GNT:-)"
freq	4825	30	3	2	2

```
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   v1           5572 non-null   object
1   v2           5572 non-null   object
2   Unnamed: 2   50 non-null     object
3   Unnamed: 3   12 non-null     object
4   Unnamed: 4   6 non-null      object
dtypes: object(5)
```

در این قسمت میخواهیم یک توزیعی از برچسب داده ها ارائه دهیم:



```
count    5572
unique      2
top       ham
freq     4825
Name: v1, dtype: object
```

همانطور که مشخص است، برچسب ها شامل دو گروه spam و ham میباشند که توزیع آن ها آورده شده است و همچنین تعداد آن ها نیز به ترتیب ۴۸۲۵ و ۷۴۷ میباشد.

۲-۷. قسمت ۲

از تابع CountVectorizer استفاده میکنیم و از ستون پیامک، ویژگی ها را استخراج میکنیم و در نهایت آن ها را به داده های آموزشی و تست تقسیم میکنیم. یک مدل SVM با این داده ها ایجاد کرده و با سرچ بهترین پارامتر بر روی مدل آموزش میدهم.

Grid search و random search دو تکنیک بهینه سازی فرایارامتر رایج در یادگیری ماشین هستند. هدف هر دو روش یافتن بهترین مجموعه فرایارامترها برای یک مدل یادگیری ماشینی است، اما آنها در رویکردشان برای کاوش فضای فرایارامتر متفاوت هستند.

۲-۷. grid search

جستجوی شبکه ای یک روش سیستماتیک است که به طور جامع از طریق مجموعه ای از مقادیر از پیش تعریف شده از فرایارامتر جستجو می کند. شبکه ای از تمام ترکیبات ممکن از مقادیر فرایارامتر ایجاد می کند و عملکرد مدل را برای هر ترکیب ارزیابی می کند.

مزایا:

- جستجوی جامع تضمین می کند که تمام ترکیب های ممکن امتحان شده است.
- اجرای آن ساده و آسان است.

معایب:

- این می تواند از نظر محاسباتی گران باشد، به خصوص زمانی که با فضای فرایارامتری بزرگ سروکار داریم.
- زمانی که تنها یک زیرمجموعه کوچک از فرایارامترها به طور قابل توجهی بر عملکرد مدل تأثیر می گذارد، ممکن است کارآمد نباشد

ما برای استفاده از grid search مجموعه پارامتر های زیر را تعریف کرده ایم:

```
param_grid = {  
    'kernel': ['linear', 'rbf'],  
    'C': [0.01, 0.1, 1, 5, 8, 10, 12, 15],  
    'gamma': [0.01, 0.1, 1, 5, 8, 10, 12, 15]  
}
```

و در نهایت بهترین پارامتر ها بصورت زیر بدست آمده است:

```
Best hyperparameters: {'C': 1, 'gamma': 0.01, 'kernel': 'linear'}
```

همچنین با تست بر روی داده های تست میزان accuracy و ماتریس پیچیدگی بصورت زیر بدست می آید:

```
Accuracy: 0.9832535885167464  
confusion matrix is:  
[[1443   1]  
 [  27 201]]
```

۷-۳. random search

از سوی دیگر، جستجوی تصادفی، مقادیر فراپارامتر را به صورت تصادفی از یک محدوده یا توزیع از پیش تعریف شده نمونه برداری می کند. از یک شبکه سیستماتیک پیروی نمی کند، و به آن اجازه می دهد فضای فراپارامتر را در برخی موارد کارآمدتر کشف کند.

مزایا:

- می تواند در مقایسه با جستجوی شبکه ای از نظر محاسباتی کارآمدتر باشد، به خصوص زمانی که تنها زیر مجموعه کوچکی از فراپارامترها تأثیرگذار هستند.
- به احتمال زیاد مناطقی از فضای فراپارامتر را که به عملکرد مدل مرتبط است، کاوش می کند.

معایب:

- هیچ تضمینی برای جستجوی جامع فضای فراپارامتر وجود ندارد.
- به دلیل ماهیت تصادفی جستجو، نتایج می توانند متغیرتر باشند.

ما برای استفاده از random search مجموعه پارامتر های زیر را تعریف کرده ایم که بصورت یکنواخت بین ۰ و ۱۵ انتخاب میشوند:

```
param_grid = {  
    'C': uniform(loc=0, scale=15),  
    'kernel': ['linear', 'rbf'],  
    'gamma': uniform(loc=0, scale=15)  
}
```

و در نهایت بهترین پارامتر ها بصورت زیر بدست آمده است:

```
Best hyperparameters: {'C': 10.229652849670448, 'gamma': 5.757482419794081, 'kernel': 'linear'}
```

همچنین با تست بر روی داده های تست میزان accuracy و ماتریس پیچیدگی بصورت زیر بدست می آید:

```
Accuracy: 0.9832535885167464  
confusion matrix is:  
[[1443   1]  
 [  27 201]]
```

همانطور که مشاهده میشود، میزان accuracy برای هر دو پارامتر بدست آمده، یکسان است و تنها وجه اشتراک آن ها نوع kernel = linear میباشد.

جستجوی شبکه ای به طور سیستماتیک تمام ترکیب های فراپارامتر از پیش تعریف شده را بررسی می کند و پوشش کامل را تضمین می کند اما می تواند از نظر محاسباتی گران باشد. در مقابل، جستجوی تصادفی مقادیر فراپارامتر را به طور تصادفی نمونه برداری می کند و کارایی را در کاوش مناطق مرتبط با هزینه محاسباتی کمتر ارائه می دهد. انتخاب به اندازه فضای فراپارامتر و منابع محاسباتی بستگی دارد، با یک جستجوی شبکه تصادفی که اغلب تعادلی بین جامع بودن و کارایی ایجاد می کند.

پاسخ ۸ – شبیه سازی

در این سوال، ابتدا دیتاست را لود کرده و اطلاعات آن را که شامل ۸۱ کلاس است را مشاهده میکنیم.

```
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Id                   1460 non-null   int64   
1   MSSubClass           1460 non-null   int64   
2   MSZoning             1460 non-null   object   
3   LotFrontage         1201 non-null   float64  
4   LotArea             1460 non-null   int64   
5   Street              1460 non-null   object   
6   Alley               91 non-null     object   
7   LotShape            1460 non-null   object   
8   LandContour         1460 non-null   object   
9   Utilities           1460 non-null   object   
10  LotConfig           1460 non-null   object   
11  LandSlope           1460 non-null   object   
12  Neighborhood         1460 non-null   object   
13  Condition1          1460 non-null   object   
14  Condition2          1460 non-null   object   
15  BldgType            1460 non-null   object   
16  HouseStyle          1460 non-null   object   
17  OverallQual          1460 non-null   int64   
18  OverallCond          1460 non-null   int64   
19  YearBuilt           1460 non-null   int64   
...
79  SaleCondition        1460 non-null   object   
80  SalePrice            1460 non-null   int64   
dtypes: float64(3), int64(35), object(43)
```

همانطور که در تصویر زیر مشخص است، تعدادی از داده ها مقدار ندارند و برابر null میباشند. بنابراین ابتدا باید این داده ها را با یک مقدار مناسب جایگزین کرد.

```
Id                0
MSSubClass        0
MSZoning          0
LotFrontage      259
LotArea          0
...
MoSold           0
YrSold           0
SaleType         0
SaleCondition    0
SalePrice        0
length: 81, dtype: int64
```

برای ستون هایی که شامل عدد هستند، داده های از دست رفته را با مقدار میانگین جایگذاری میکنیم. و ستون هایی که را که از نصف دیتاست داده کمتر دارند را نیز حذف میکنیم.

در نهایت تعداد کلاس های بدست آمده برابر با ۳۸ کلاس خواهد بود که ما مدل خود را بر اساس این کلاس ها آموزش میدهیم. البته با توجه به بالا بودن تعداد ویژگی ها، ابتدا با SelectKBest بهترین ویژگی ها را براساس f_regression انتخاب میکنیم، ما اینجا مقدار k را برابر ۲۵ قرار داده ایم.

ویژگی های انتخابی به شرح زیر میباشد:

```
['LotFrontage', 'LotArea', 'OverallQual', 'YearBuilt', 'YearRemodAdd',  
'MasVnrArea', 'BsmtFinSF1', 'BsmtUnfSF', 'TotalBsmtSF', '1stFlrSF',  
'2ndFlrSF', 'GrLivArea', 'BsmtFullBath', 'FullBath', 'HalfBath',  
'BedroomAbvGr', 'KitchenAbvGr', 'TotRmsAbvGrd', 'Fireplaces', 'GarageYrBlt',  
'GarageCars', 'GarageArea', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch']
```

حال دیتا ها را نرمالایز کرده و سپس یک مدل Support vector regression را بر روی آن لحاظ میکنیم.

ما در این جا دو نوع کرنل را تست کردیم و نتایج آن ها به صورت زیر خواهد بود:

Linear kernel:

Train MSE: 0.155

Validation MSE: 0.285

Test MSE: 0.416

RBF kernel:

Train MSE: 0.092

Validation MSE: 0.288

Test MSE: 0.144

نتایج کرنل RBF به دلیل پیچیدگی و افزایش بیشتر بعد فضای داده ها، بهتر است.

با افزایش تعداد ویژگی ها، مدل دقیق تر میشود اما با توجه به اینکه داده ها نسبت به پارامتر مدل کم است، بنابراین امکان بیش برآزش وجود دارد.

ما بصورت آزمایشی تعداد داده های متفاوت را در نظر گرفتیم و بهترین تعداد را انتخاب کردیم.

از طرفی ممکن است، برخی از ویژگی ها با یکدیگر دارای همبستگی باشند که اضافه نمودن آن ها به فضای ویژگی منجر به افزایش پارامتر ها و در نهایت منجر به افزایش پیچیدگی مدل میشود.