

## تمرین هفته یازدهم بوتکمپ یادگیری ماشین

### ۱. مسئله رگرسیون:

بر روی مجموعه داده مشخص شده ، حداقل ۴ مدل مختلف را آزمایش نمایید و بر اساس **خطای میانگین مربعات** بهترین مدل را بیابید. در شکل زیر ویژگی‌های مختلف این مجموعه داده نشان داده شده است. شما می‌توانید به صورت شهودی و یا با استفاده از تکنیک‌های استخراج ویژگی بهترین ویژگی‌ها را بیابید.

**توضیحات در رابطه با مجموعه داده:** این مجموعه داده شامل اطلاعات مربوط به اجاره دوچرخه در سال‌های ۲۰۱۱ و ۲۰۱۲ در سامانه capital bikeshare می‌باشد که در برخی از شهرهای آمریکا مورد استفاده قرار می‌گیرد. تعداد دوچرخه‌های اجاره شده به شدت به شرایط آب و هوایی و برخی ویژگی‌های دیگر که در زیر بیان شده است بستگی دارد. ویژگی‌هایی که در این مجموعه داده وجود دارد در شکل زیر نشان داده شده است. بر این اساس شما باید تعداد دوچرخه‌های اجاره شده را با توجه به سایر ویژگی‌ها تخمین بزنید. این مجموعه داده دو حالت دارد که شما مختار به استفاده از هر کدام از آن‌ها می‌باشید. در حالت اول که با عنوان "day.csv" مشخص شده است، تعداد دوچرخه‌ها در روزهای پیاپی گردآوری شده‌اند و بررسی در سطح روزانه می‌باشد. این حالت دارای ۷۳۱ رکورد می‌باشد. در حالت دوم تعداد دوچرخه‌ها در ساعات پیاپی مشخص شده‌اند و دارای ۱۷۳۷۹ رکورد می‌باشد.

## Data Dictionary

Column Position	Attribute Name	Definition	Data Type	Example	% Null Ratios
1	instant	Record Index	Quantitative	190, 7, 17180	0
2	dteday	Date (Format: YYYY-MM-DD)	Quantitative	2012-12-23, 2012-01-01, 2012-06-24	0
3	season	Season (1: springer, 2: summer, 3: fall, 4: winter)	Quantitative	1, 2, 4	0
4	yr	Year (0: 2011, 1:2012)	Quantitative	0, 1	0
5	mnth	Month (1 to 12)	Quantitative	1, 6, 12	0
6	hr	Hour (0 to 23) - Not in day.csv dataset	Quantitative	4, 6, 14	0
7	holiday	Weather day is holiday or not	Quantitative	0, 1	0
8	weekday	Day of the week	Quantitative	0, 6, 3	0
9	workingday	Working Day: If day is neither weekend nor holiday is 1, otherwise is 0	Quantitative	0, 1	0
10	weathersit	Weather Situation (1: Clear, Few clouds, Partly cloudy, Partly cloudy; 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist; 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds, 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog)	Quantitative	1, 2, 3	0
11	temp	Normalized temperature in Celsius. The values are derived via $(t - t_{min}) / (t_{max} - t_{min})$ , $t_{min} = -8$ , $t_{max} = +39$ (only in hourly scale)	Quantitative	0.08, 0.22, 0.34	0
12	atemp	Normalized feeling temperature in Celsius. The values are derived via $(t - t_{min}) / (t_{max} - t_{min})$ , $t_{min} = -16$ , $t_{max} = +50$ (only in hourly scale)	Quantitative	0.0909, 0.2727, 0.303	0
13	hum	Normalized humidity. The values are divided to 100 (max)	Quantitative	0.53, 0.8, 0.31	0
14	windspeed	Normalized wind speed. The values are divided to 67 (max)	Quantitative	0.194, 0, 0.2985	0
15	casual	Count of casual users	Quantitative	0, 2, 57	0
16	registered	Count of registered users	Quantitative	1, 0, 118	0
17	cnt	Count of total rental bikes including both casual and registered	Quantitative	1, 2, 175	0

## ۲. مسئله کلاس‌بندی:

بر روی مجموعه داده مشخص شده زیر، حداقل ۴ الگوریتم مختلف که هر کدام شامل مدل‌هایی با پارامترهای متفاوت است آزمایش نمایید و بر اساس معیار دقت ۱۰ بهترین مدل را بیابید. نمایش ماتریس پیرشانی برای این مجموعه داده الزامی می‌باشد. در شکل زیر ویژگی‌های مختلف این مجموعه داده نشان داده شده است. شما می‌توانید به صورت شهودی و یا با استفاده از تکنیک‌های استخراج ویژگی بهترین ویژگی‌ها را بیابید.

**توضیحات در رابطه با مجموعه داده:** این مجموعه داده شامل اطلاعات ۱۵۵ بیمار می‌باشد که مشکوک به بیماری هپاتیت می‌باشند. هر کدام از این رکوردها حداکثر دارای ۲۰ ویژگی می‌باشند (برخی از رکوردها دارای مقادیر گم شده می‌باشد که رویکرد شما در برخورد با این مقادیر بر عهده خودتان می‌باشد). در این مجموعه داده شما باید این بیمارها را بر اساس ویژگی‌های هر کدام به دو کلاس live یا die کلاس‌بندی نمایید.

## Data Dictionary

Column Position	Attribute Name	Definition	Data Type	Example	% Null Ratios
1	Class	Class (1: DIE, 2: LIVE)	Quantitative	1, 2	0
2	Age	Age (In Years)	Quantitative	34, 20, 55	0
3	Sex	Sex (1: Male, 2: Female)	Quantitative	1, 2	0
4	Steroid	Steroid (No: 1, Yes: 2)	Quantitative	1, 2	1
5	Antivirals	Antivirals (No: 1, Yes: 2)	Quantitative	1, 2	0
6	Fatigue	Fatigue (No: 1, Yes: 2)	Quantitative	1, 2	1
7	Malaise	Malaise (No: 1, Yes: 2)	Quantitative	1, 2	1
8	Anorexia	Anorexia (No: 1, Yes: 2)	Quantitative	1, 2	1
9	Liver Big	Liver Big (No: 1, Yes: 2)	Quantitative	1, 2	10
10	Liver Firm	Liver Firm (No: 1, Yes: 2)	Quantitative	1, 2	11
11	Spleen Palpable	Spleen Palpable (No: 1, Yes: 2)	Quantitative	1, 2	5
12	Spiders	Spiders (No: 1, Yes: 2)	Quantitative	1, 2	5
13	Ascites	Ascites (No: 1, Yes: 2)	Quantitative	1, 2	5
14	Varices	Varices (No: 1, Yes: 2)	Quantitative	1, 2	5
15	Bilirubin	Bilirubin	Quantitative	0.39, 0.80, 1.20	6
16	Alk Phosphate	Alk Phosphate	Quantitative	33, 80, 120	29
17	Sgot	SGOT	Quantitative	13, 100, 200	4
18	Albumin	Albumin	Quantitative	2.1, 3.0, 3.8	16
19	Protime	Protime	Quantitative	60, 70, 80	67
20	Histology	Histology (No: 1, Yes: 2)	Quantitative	1, 2	0