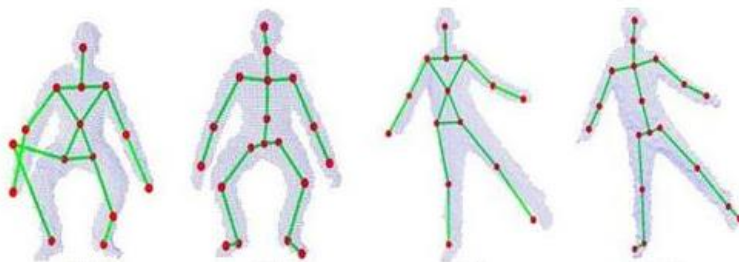


بسمه تعالی



دانشگاه صنعتی شریف  
دانشکده مهندسی برق  
گروه سیستم‌های دیجیتال



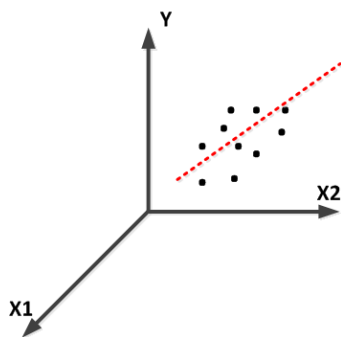
## آزمایشگاه یادگیری و بینایی ماشین

دستور کار آزمایش دوم: رگرسیون خطی

زمان لازم برای انجام آزمایش: حداکثر دو جلسه

### 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 پیش بینی 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، نمونه ای از آلودگی هوا را بگیریم، می توانیم نمودار زیر را تشکیل دهیم.



همان طور که دیده می شود، تا وقتی که داده جمع آوری شده، کاملاً نویز نباشد، می توان 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 خط چین نمایش داده ایم.

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

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

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

$$y = f(x) + \epsilon$$

که در آن،  $\epsilon$  نویز است. هدف ما یافتن  $f(\cdot)$  است که همان رگرسیون خطی داده است:

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

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

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

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

برای یافتن  $f$  ما باید ضرایب  $\beta$  را پیدا کنیم. باید  $\beta$  هایی را بیابیم که کمترین اختلاف بین  $y^i$  ها و  $f(x^i)$  ها را ایجاد کند تا بهترین تخمین ممکن را داشته باشیم. ما برای نمایش تفاضل این دو مقدار، از تفاضل نرم  $\ell_2$  استفاده می‌کنیم:

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

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

$$RSS = (y - X\beta)^T (y - X\beta)$$

که در آن، ماتریس  $X$  به صورت زیر است:

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

و  $y$  بردار ستونی است که خروجی نمونه‌ها را نشان می‌دهد.

(امتیازی): رابطه ماتریسی RSS را اثبات نمایید.



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

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

از رابطه بالا برای تخمین  $y$  (خروجی نمونه‌های آموزشی) به صورت زیر استفاده می‌شود:

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



ابتدا داده های دستورکار را که نمایانگر میزان ازن موجود در هوا براساس میزان تابش خورشید، دما و سرعت وزش باد می باشد را بخوانید سپس ۸۰ درصد داده ها را به صورت رندم به عنوان داده آموزش و مابقی را به عنوان داده تست جدا کنید. درنهایت ضرایب رگرسیون را با استفاده از مجموعه آموزش بدست آورید. میانگین توان دو خطا را هم روی داده های تست و هم روی داده های آموزش بدست آورید.



اکنون، داده های میزان ازن هوا را با نویز گوسی با واریانس های گوناگون در بازه [0,50] نویزی نمایید. سپس داده ها هر بار به آموزش و تست تقسیم کرده و رگرسیون خطی انجام دهید. اثر تغییرات واریانس نویز را بر روی خطای داده های آموزش و تست بر روی یک نمودار نشان دهید و تحلیل نمایید.



توان دو متغیرهای موجود را به ماتریس  $X$  اضافه نمایید و مجدد خطاهای آموزش و تست را محاسبه نمایید.

### قسمت دوم: رگرسیون خطی با تعداد ویژگی های بالا

در این قسمت می خواهیم شدت احساس تعجب را با استفاده از تصویر چهره فرد تخمین بزنیم. برای آموزش و تست رگرسور از مجموعه داده surprise video استفاده کنید. در این مجموعه فریم های مربوط به ویدیوی ضبط شده از چهره ۳۰ فرد از شروع از حالت خنثی تا اوج تعجب وجود دارد. با استفاده از یک فرض ساده کننده، شدت احساس را به صورت خطی از مقدار صفر، برای حالت خنثی در فریم اول، تا یک، برای اوج احساس در فریم آخر، برچسب بزنید. ۶۰ درصد از ویدیو ها را به صورت رندم برای آموزش و مابقی را برای تست کنار بگذارید.



حال با استفاده از مقادیر پیکسلها به عنوان ورودی و برچسبها به عنوان خروجی، یک رگرسور را آموزش دهید. آیا آموزش رگرسور با این تعداد ویژگی امکانپذیر است؟



همان طور که در فریمها قابل مشاهده است، می توانید با استفاده از یک ماسک مناسب قسمت‌های اضافی تصاویر را حذف کنید تا تعداد پیکسلها به طور قابل ملاحظه ای کاهش یابد. علاوه بر این، با تغییر اندازه طول و عرض فریمها تا مثلاً یک چهارم اندازه اولیه، تعداد پیکسلها را می توانید باز هم کاهش دهید. تلاش برای آموزش رگرسور را تکرار کنید. اگر تلاش شما موفقیت آمیز بود، برچسبهای فریمهای افراد آموزش و تست را تخمین بزنید و خطاهای آموزش و تست را بدست آورید. از مقایسه خطای آموزش با خطای تست چه نتیجه ای می گیرید؟



از روش PCA که در آزمایش اول آموختید برای کاهش مؤثرتر بعد داده ها استفاده کنید. دقت کنید که باید فضای PCA را با استفاده از داده های آموزش بسازید و سپس هم داده های آموزش و هم داده های تست را با همان فضا کاهش بعد بدهید. بعد از آموزش رگرسور، برچسبهای فریمهای افراد آموزش و تست را تخمین بزنید و خطای رگرسیون در آموزش و تست را محاسبه کنید. خطای آموزش و خطای تست را با قسمت قبل (استفاده از پیکسلهای خام به عنوان ویژگی) مقایسه کنید و تحلیل خود را بنویسید. بهترین خطای تست را با چه تعداد ویژگی بدست می آورید؟

## پیش گزارش

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

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

$\beta_0^*$  و  $\beta_1^*$  جواب‌های این مسأله هستند. از بین معادله‌های زیر، آن‌هایی که صحیح هستند را مشخص کنید.  
(راهنمایی: از معادله بالا نسبت به  $\beta_0$  و  $\beta_1$  مشتق بگیرید.)

- $\sum_{i=1}^n (y^i - \beta_0^* - \beta_1^* x^i) y^i = 0$
- $\sum_{i=1}^n (y^i - \beta_0^* - \beta_1^* x^i) (y^i - \bar{y}) = 0$
- $\sum_{i=1}^n (y^i - \beta_0^* - \beta_1^* x^i) (x^i - \bar{x}) = 0$
- $\sum_{i=1}^n (y^i - \beta_0^* - \beta_1^* x^i) (\beta_0^* + \beta_1^* x^i) = 0$

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