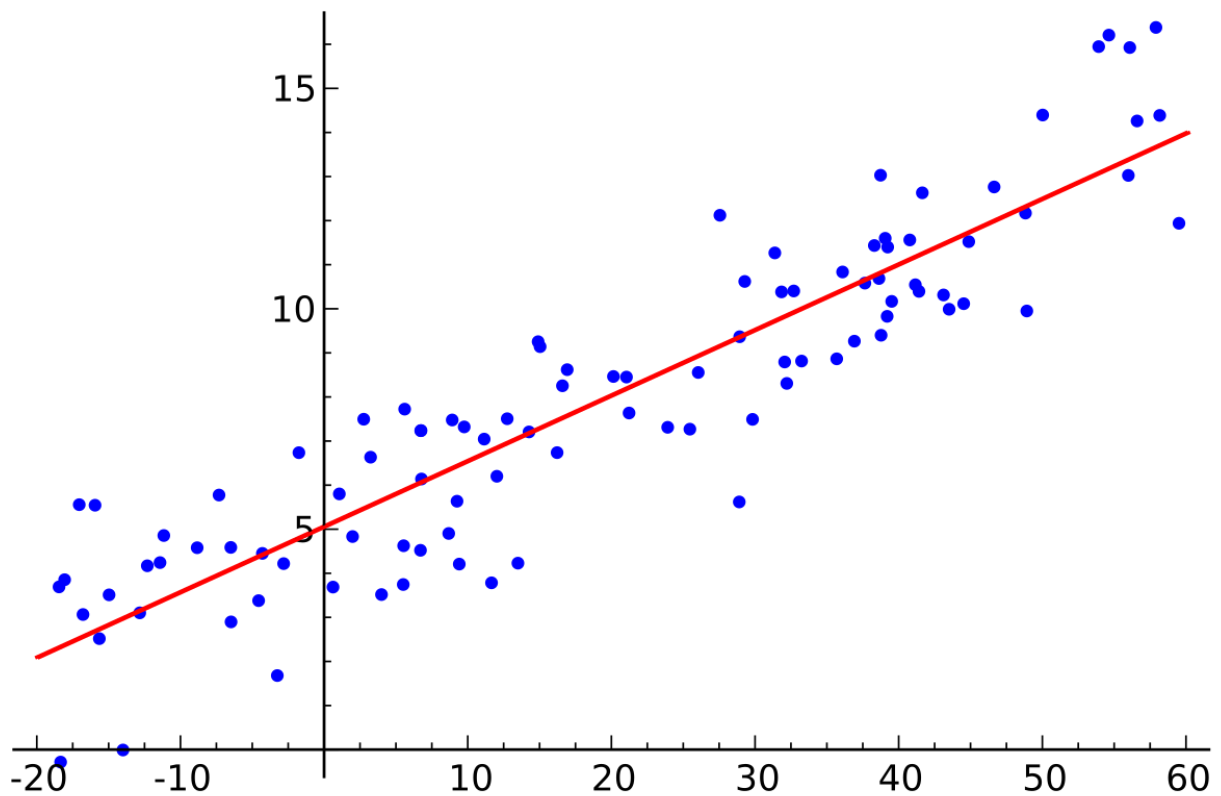


به نام خدا

## تکلیف جلسه سوم آزمون برنامه سازی پیشرفته – رگرسیون خطی

در این جلسه قصد داریم یک مدل یادگیری ماشین مبتنی بر رگرسیون خطی را پیاده سازی کنیم. این مدل باید توانایی این را داشته باشد که با استفاده از یک مجموعه داده آموزشی تعلیم داده شود و پیش بینی درستی مبتنی بر مدل رگرسیون خطی به ما ارائه دهد.

ابتدا با رگرسیون خطی آشنا می شویم. به این شکل توجه کنید:



فرض کنید داده های  $y$ ، داده های آموزشی باشند که برای تعلیم سیستم از آنها استفاده کرده ایم. حال می خواهیم به ازای یک  $x$  دلخواه طبق این داده ها، مقدار  $y$  را پیش بینی کنیم. برای این کار یک خط را بر روی این داده ها قرار می دهیم به شکلی که کمترین میزان خطا را به ما بدهد. در این صورت با کمک معادله این خط می توانیم پیش بینی را انجام دهیم.

اگر معادله خط به فرم  $y = \hat{\beta}_0 + \hat{\beta}_1 x$  باشد، پارامترها بدین شکل تخمین زده می شوند:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

که در آن  $\bar{x}$  و  $\bar{y}$  به ترتیب میانگین  $x$  ها و  $y$  ها و  $N$  تعداد داده ها می باشند.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

خطای این مدل بدین شکل بدست می آید:  $Error = \frac{1}{N} \sum_{i=1}^N e_i^2$  که  $e_i$  ها اختلاف  $y$  هر داده با مقدار پیش بینی شده می باشد.

حال به پیاده سازی این مدل می پردازیم:

ابتدا کلاسی طراحی می کنیم به اسم **pair** که وظیفه دارد یک زوج مرتب **x** و **y** را نگه دارد. اجزای داده این کلاس عبارت اند از:

- دو داده **float** به نام های **x** و **y**
- یک عدد صحیح استاتیک برای شمردن تعداد زوج مرتب ها

توابع این کلاس عبارت اند از:

- یک سازنده پیش فرض که **x** و **y** را ۰ می گذارد.
- یک سازنده پارامترپذیر که این دو مقدار را به صورت پارامتر دریافت می کند.
- دو تابع **getter** برای **x** و **y**
- یک تابع استاتیک برای گرفتن عضو استاتیک
- یک مخرب

برای این که تعداد داده های آموزشی ما معلوم نیست از یک لیست پیوندی برای نگه داری آن ها استفاده کنید. یک کلاس لینک لیست طراحی کنید که پویتری از جنس **pair** را به عنوان ابتدای لیست ذخیره کند. همچنین این کلاس باید شامل متدهای زیر باشد:

- یک سازنده پیش فرض
- یک تابع برای اضافه کردن یک نود به انتهای آن. لطفا این تابع را به گونه ای پیاده سازی کنید که امکان فراخوانی آبشاری آن به وجود بیاید. یعنی به عنوان مثال اگر نام شیء لینک لیست شما **list** باشد و نام تابع شما **insert** باشد، برای اضافه کردن سه زوج **p1** و **p2** و **p3** کدنویس بتواند این گونه بنویسد: **list->insert(p1)->insert(p2)->insert(p3)** (راهنمایی: برای این کار بایستی در انتهای اضافه کردن پویتر و یا رفرنسی به شیء کنونی را برگردانید – راهنمایی بیشتر!!!): از پویتر **this** کمک بگیرید)
- یک تابع میانگین گیری که یک **pair** برگرداند که **x** آن میانگین **x** ها و **y** آن میانگین **y** ها باشد.
- تابعی که با گرفتن پارامترهای خط به عنوان آرگومان ورودی، خطای مدل را بدست آورد.
- مخربی که لیست پیوندی را نابود کرده و حافظه اشغال شده را به سیستم عامل بازگرداند.

در نهایت یک کلاس برای مدل بنویسید. این کلاس داده های زیر را نگه می دارد:

- یک لینک لیست به عنوان داده های آموزشی
- پارامترهای خط رگرسیون

متدهای این کلاس را بدین شکل پیاده سازی کنید:

- در سازنده این کلاس آدرس یک فایل را دریافت کنید که در آن داده های آموزشی نگه داری می شوند. سپس این داده ها را به لینک لیست اضافه کنید و پارامترها را محاسبه و تنظیم کنید. فرمت فایل ورودی به این شکل است که در هر خط یک زوج مرتب نوشته می شوند که میان **x** و **y** یک تب فاصله وجود دارد. (نمونه آن به پیوست ارسال می شود)
- یک تابع **predict** بنویسید که با دریافت یک **x**، **y** پیش بینی شده توسط مدل را به ما برگرداند.
- یک تابع به اسم **info** بنویسید که مشخصات مدل اعم از تعداد داده آموزشی، معادله خط رگرسیون مدل و همچنین خطای مدل را چاپ کند.

برای پیاده سازی به این نکات توجه کنید:

- باید این کلاس ها را با رعایت کامل اصول شیء گرایی (به خصوص اصل data hiding) طراحی کنید.
- اصولی و بهینه کد زدن نیز فراموش نشود.
- تمام توابعی که قابلیت تعریف به صورت تابع ثابت را دارند را ثابت تعریف کنید.
- لطفا به صورت چند فایل (به ازای هر کلاس یک فایل هدر و یک فایل سورس و همچنین یک فایل برای main) این تکلیف را پیاده سازی کنید.