**Multiple Linear Regression**

برای داده های کیفی از dummy varianle استفاده میکنیم  
 dummy variable trap

استفاده از داده های بیش از اندازه دامی در مدل زیرا به عنوان مثال وقتی دو داده دامی داریم یکی از آن ها کافی است زیرا داده دوم در مقدار ثابت معادله لحاظ شده است

وقتی n داده دامی در مدل استفاده می شود باید n-1 متغیر در مدل استفاده شود

‌ساختن مدل

حذف متغیر های بی تاثیر به دو دلیل :



روش های ساختن مدل :

1. All in
2. Backward elimination
3. Froward Selection
4. Bidirectional Elimination
5. Score Comparison

All in

چه زمانی :

1. اطلاع قبلی از اینکه همه متغییر ها تاثیر دارند
2. وقتی مجبورید از همه استفاده کنید
3. امادگی برای Backward Elimination

Backward elimination

مراحل :

۱- انتخاب سطح معنی داری (significance level) در اینجا ۵٪ را انتخاب میکنیم (SL= 0.05)

۲- وارد کردن تمام متغیر ها در مدل و اجرای مدل

۳- بزرگترین مقدار p برای برآورد گر ها (متغیر های مستقل) در مدل رو انتخاب میکنیم :

برای مدل چند متغیره مسئله رابطه سود دهی استارتاپ ها و هزینه تحقیق و توسعه، هزینه مدیریت، هزینه بازاریابی و ایالتی که در آن فعالیت می کنند به این صورت هستند برآورد گر ها :

ایالت به صورت متغییر دامی وارد شده و با D نمایش داده شده در مدل

* اگر p از سطح معنی داری که انتخاب کردیم بزرگتر بود میریم مرحله ۴ اگر نبود میریم به مرحله پایانی

۴- در این مرحله متغیری که بیشترین  مقدار p را دارد حذف میکنیم

۵- مجددا مدل را برای متغییر های باقی مانده پردازش میکنیم

بعد از این مرحله باز میگردیم به مرحله ۳ تا جایی که مقدار p از سطح معنی داری که انتخاب کردیم بزرگتر نباشد در این صورت به مرحله پایانی میرویم

مرحله پایانی یعنی مدل آماده است

Forward Selection

مراحل :

۱- انتخاب سطح معنی داری (significance level) در اینجا ۵٪ را انتخاب میکنیم (SL= 0.05)

۲- برای تمامی متغیر ها یک مدل رگرسیون ساده اجرا میکنیم یعنی برای n متغیر n مدل رگرسیون ساده داریم حالا از بین این مدل ها مدلی که کمترین مقدار p برای متغیر را دارد انتخاب میکنیم

۳- حالا یک مدل رگرسیونی با دو متغیر یکی آن که انتخاب کردیم و دیگری سایر متغیر ها اجرا میکنیم یعنی الان n-1 مدل رگرسیون دو متغیره داریم

۴- از میان متغیر های دومی که اضافه کردیم و مدل رگرسیون دو متغیره تشکیل دادیم آن که کمترین مقدار p را دارد انتخاب میکنیم (مقدار p برای متغیر دومی که در مرحله ۳ اضافه کردیم) اگر مقدار p از سطح معنی داری ما کمتر بود باز به مرحله سه می رویم و حالا رگرسیون ۳ متغیره با دو متغیر انتخابی و سایر متغیر ها تشکیل میدهیم. اگر مقدار p از سطح معنی داری ما کمتر نبود به مرحله آخر می رویم

۵- در این مرحله مدلی که در آخر ساختیم را انتخاب نمیکنیم و مدل قبل آن را انتخاب میکنیم زیرا مدل آخر دارای متغیری است که در سطح معنی داری ما قرار نمیگیرد( p کمتر از سطح معنی داری ما نیست) ولی مدل قبلی قرار میگیرد

Bidirectional Elimination ( Stepwise Regression)

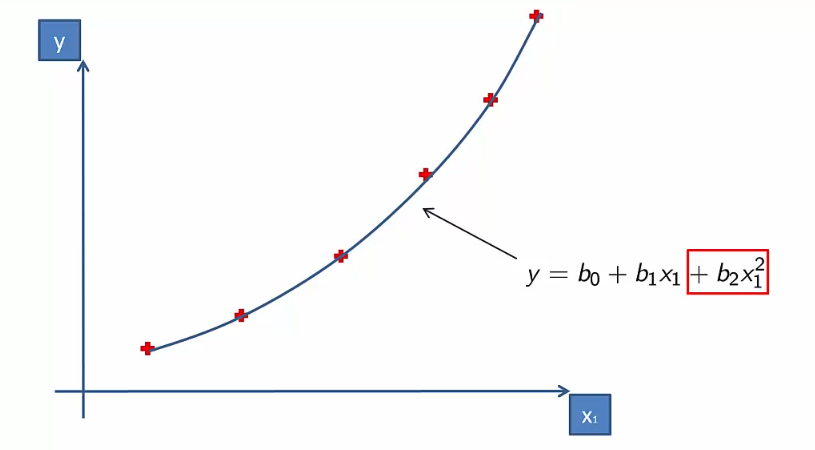
مراحل :

۱-  دو سطح معنی داری (significance level) برای ورود و ماندن انتخاب میکنیم. در اینجا ۵٪ را برای هر دو انتخاب میکنیم (SLenter= 0.05 , SLstay = 0.05)

۲-

Polynomial Linear Regression

مدل به صورت زیر خواهد بود :



در این دیتا ست اگر از رگرسیون خطی ساده استفاده کنیم با جلوتر رفتن داده ها از خط رگرسیون فاصله میگیرند برای حل این مشکل از رگرسیون نمایی استفاده میکنیم

چرا همچنان رگرسیون خطی است؟

چون بر اساس ضرایب مدل تعیین میشود و ضرایب هستند که مجهولند و ضرایب همچنان معادله خطی با یکدیگر دارند

مثال استفاده : این مدل برای شیوع بیماری ها استفاده می شود