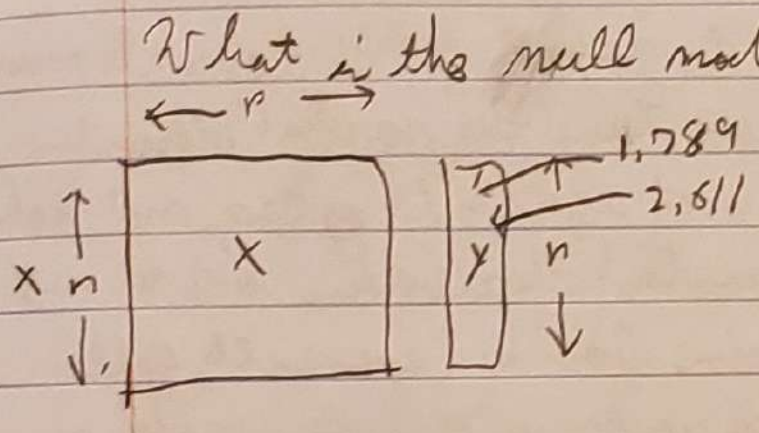
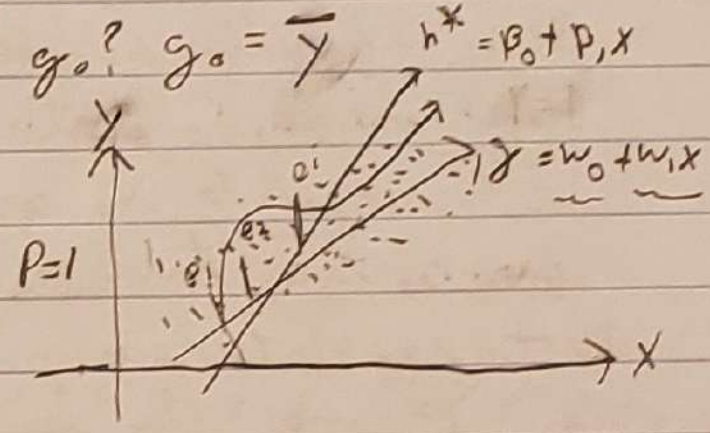
2/17/21

So far, the response space was $\{0,1\}$ and the models were "binary classification" models. What if $y = \mathbb{R}$ or $y \subset \mathbb{R}$? This means the response is continuous and our prediction will be continuous. These models are called "regression" models. The word "regression" is used because of historical circumstance only (see lab).

What is the null model g_0? $g_0 = \bar{y}$





$$\mathcal{H} = \{ \vec{w} \cdot \vec{x} ; \vec{w} \in \mathbb{R}^{p+1} \}$$
$$w_0 + w_1 x_1 + \ldots w_p x_p$$

Like before, this candidate set, requires a "1" appended to each of the original p-length x-vectors.

$$h^*(\vec{x}) = w_0^* + w_1^* x_1 + \ldots w_p^* x_p = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

$$\downarrow$$

$$y = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p + \varepsilon$$

Standard notation for the best / "true" value of the linear coefficients.

We have training data and the candidate set of linear models. We need an algorithm that will compute w_0 and w_1 for us. We first need an "objective function" or "error function" or "loss function" which gives the degree of our model mistakes

Let $e_i := y_i - \hat{y}_i$. Consider the loss function:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - w_0 - w_1 x_i)^2$$

(sum of squared error)

Our algorithm will seek to argmin {SSE} over all possible w_0, w_1 values. To do this, we take the partial derivative with respect to w_0 and set equal to zero and solve for b_0 then take the partial derivative w.r.t w_1 and set equal to zero and solve for b_1. we will call $g(x) = b_0 + b_1 x$ the "least squares" regression model or "ordinary least squares" (OLS)

$$\sum y_i^2 + w_0^2 + w_1^2 x_i^2 - 2y_i w_0 - 2y_i w_1 x_i + 2w_0 w_1 x_0$$

$$= \sum y_i^2 + n w_0^2 + w_1^2 \sum x_i^2 - 2w_0 n\bar{y} - 2w_1 \sum x_i y_i + 2w_0 w_1 n\bar{x}$$

$$\frac{\partial}{\partial w_0}(SSE) = \frac{\partial}{\partial w_0}[\;] = 2nw_0 - 2n\bar{y} + 2w_1 n\bar{x} \overset{\text{set}}{=} 0 \Rightarrow b_0 =$$

$$b_0 = \frac{n\bar{y} - w_1 n\bar{x}}{n} = \bar{y} - b_1\bar{x}$$

$$\frac{\partial}{\partial w_1}[SSE] = \frac{\partial}{\partial w_1}[\;] = 2w_1 \sum x_i^2 - 2\sum x_i y_i + 2w_0 n\bar{x} \overset{\text{set}}{=} 0$$

$$\Rightarrow b_1 \sum x_i^2 = \sum x_i y_i - b_0 n\bar{x} = \sum x_i y_i - (\bar{y} - b_1\bar{x}) n\bar{x}$$

$$\Rightarrow b_1 \sum x_i^2 = \sum x_i y_i - n\bar{x}\bar{y} + n\bar{x}^2 b_1 \Rightarrow b_1 \sum x_i^2 - b_1 n\bar{x}^2 = \sum x_i y_i - n\bar{x}\bar{y}$$

$$\Rightarrow b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$ this is the answer and now we simplify it... using Math 241-like notation

$$s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - 2\bar{x}\underbrace{\sum x_i}_{n\bar{x}} - n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)$$

$$= \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$

all possible
rearrange
and solve
w. and
call
ded or

$$\rho := Corr[X,Y] := \frac{Cov[X,Y]}{SE[X]\,SE[Y]} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{Var[Y]\,Var[X]}}$$

Covariance is extended with

$$s_{xy} = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum x_i y_i - \bar{y}\underbrace{\sum x_i}_{n\bar{x}} - \bar{x}\underbrace{\sum y_i}_{n\bar{y}} + n\bar{x}\bar{y}\right)$$

$$= \frac{1}{n-1}\left(\sum x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}\right) = \frac{1}{n-1}\left(\sum x_i y_i - n\bar{x}\bar{y}\right)$$

$$b_1 = \frac{(n-1)s_{xy}}{(n-1)s_x^2} = \frac{s_{xy}}{s_x^2} \qquad r := \frac{s_{xy}}{s_x s_y} \Rightarrow s_{xy} = r\,s_x s_y$$

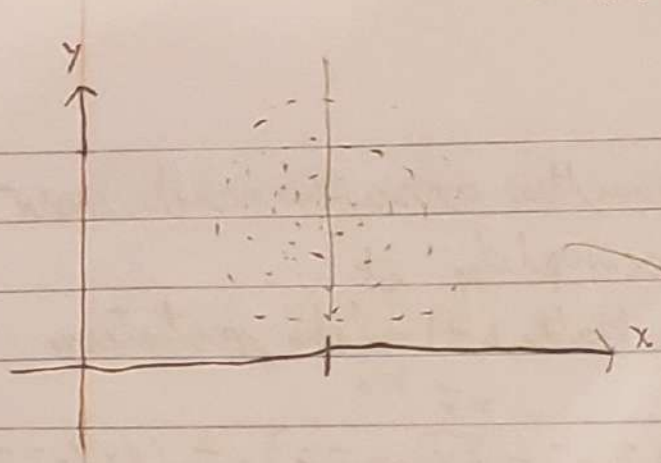
$$= \frac{r\,s_x s_y}{s_x^2} = r\frac{s_y}{s_x} \Rightarrow b_0 = \bar{y} - r\frac{s_y}{s_x}\bar{x}$$

two w, $n\bar{x}$

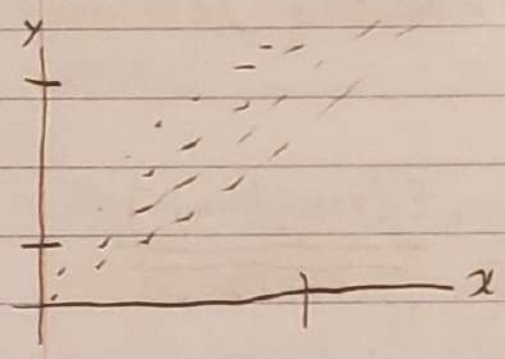
$\Rightarrow b_0 =$

$\sum x_i y_i - n\bar{x}\bar{y}$

Cor[X,Y] = 0

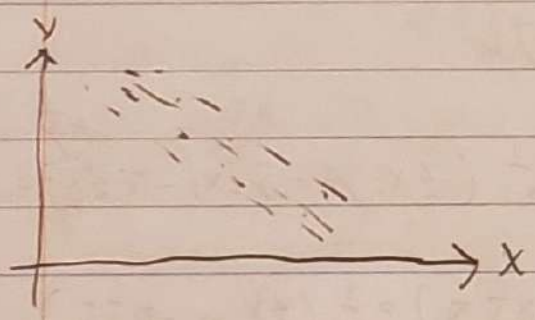


Covariance measures change in expected value of the second rv if the first rv changes

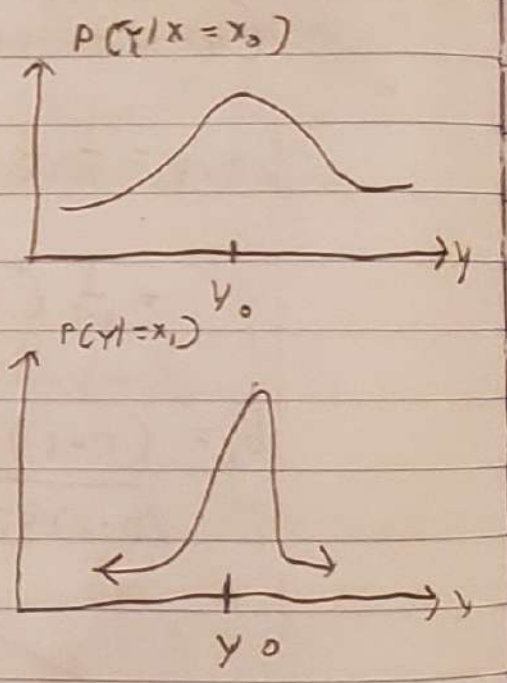


Cor[X,Y] > 0

Are X, Y independent?



Cor[X,Y] < 0



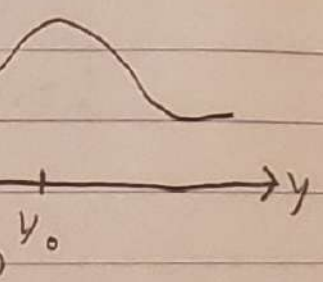
The word "association" just means "dependence". Correlation means linear dependence (and covariance mean linear dependence).

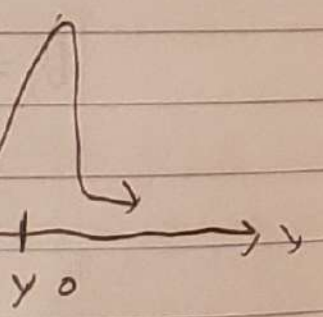
Correlation is a types of association (it is linear association).

Let examine a special case of OLS where p=1. Let the only feature be a binary feature eg. x_i is either "red" or "green". Lets create a new x which is a dummy / binary variable which is 0 if red and 1 if green. What is a good model for prediction

easures
pected value
rv it. the
ger

independent?
$= x_0$)





the only
or "green".
ble which
on prediction

$g(red) = \bar{y}_{red}$

$g(green) = \bar{y}_{green}$

} OLS model

x	y
0	3.21
1	8.43
1	6.22
0	1.07
1	7.11
⋮	7.10
⋮