Least Sa, van Gradient Descent
first lets comput gradient wit single param 0; &
Single observation (x) & then vectorize so!" to update page
Least Sq. var Gradient Descent
-> one als at a time / -> I that in / online grown have.
for single obs, error => J = (y-xe)
for single obs, error => $J = (y - x\theta)^2$ $\Rightarrow \frac{d}{d\theta} (y^2 - 2yx\theta + (x\theta)^2)$
$\frac{\partial}{\partial \theta} = 2 \left(\frac{1}{x_0 - y} \right) \frac{1}{x_0} = \frac{1}{x_0 - y} \frac{1}{x_0} \frac{1}{x_0} = \frac{1}{x_0 - y} \frac{1}{x_0} \frac{1}{x_0$
$\frac{d}{d} = \frac{1}{2(x_0 - y)} \chi_i $
$d\theta_{\rm j}$
2646 WY
$\frac{d^{J}}{d\theta} = 2x^{T}(x\theta - y)$ $= x^{T}$
$d\theta = x^{-1}$
Learning Rate: $ \frac{\partial \theta}{\partial \theta} = \theta - 2x^{\dagger} (x\theta - y) + \frac{\partial \theta}{\partial \theta} $
Learning Rate:
J do if m more params according to do
are can go too slowly - for much him
go too tast -> Ourshoot desired Minima/man
(pasically sty size)
we use in direction of a redient
(maxina stro-size Alnamic Surta)
The do if m more params according to dt are can go too slowly a four much him go too slowly are can minimal man according to desired minimal man are use n called learning rate that contols how much to go in a direction of a radient. (maxing step-size dynamic sorta) Standardize the data or n gos te crazy!!