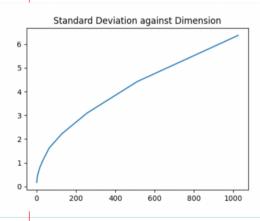
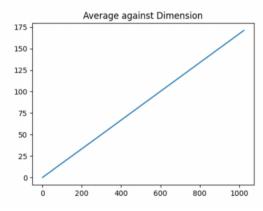
Question (a)





Questim 16) let R = Z1 + Zz + - - + Zn where Zz = (Xi - Yi)2

notive that for Ei, E; for lei, jen Zi and Z; are independent, because

the difference of distance on one dimension

is independent on another dimension

We Know E(71)=6, Var (2:1=180

by Linearity: E[R] = E[Zit - + Zi] = E[Zi] + -- + + [Zi]

=6 E[Zi] = &

by independence: Var(R] = Var[Zit --- † Zd]

= Var[2,] + --- + Var[2]

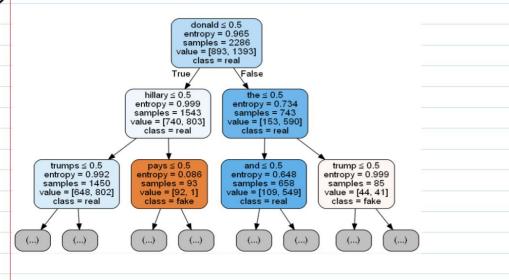
= d Var[Zi]

= 1200

## Questin Zb)

decision tree classifier using max\_depth 5 and split criteria information gain has accuracy 0.7061224489795919
decision tree classifier using max\_depth 10 and split criteria information gain has accuracy 0.7081632653061225
decision tree classifier using max\_depth 10 and split criteria information gain has accuracy 0.7081632653061225
decision tree classifier using max\_depth 10 and split criteria Gini coefficient has accuracy 0.71836734693878
decision tree classifier using max\_depth 50 and split criteria information gain has accuracy 0.7591836734693878
decision tree classifier using max\_depth 50 and split criteria Gini coefficient has accuracy 0.7489795918367347
decision tree classifier using max\_depth 100 and split criteria information gain has accuracy 0.7612244897959184
decision tree classifier using max\_depth 100 and split criteria Gini coefficient has accuracy 0.7428571428571428
decision tree classifier using max\_depth 150 and split criteria information gain has accuracy 0.77551020408016326
decision tree classifier using max\_depth 150 and split criteria Gini coefficient has accuracy 0.7673469387755102

## 24)



29)	
feature_name	s, data_train, label_train, data_valid, label_valid, data_test, label_test = load_data() (feature_names, data_train, label_train, data_valid, label_valid, data_test, label_test)
	ation gain at root split donald is {compute_information_gain(data_train, label_train, "donald", feature_names)}')
f'informa	ation gain at other keyword the is {compute_information_gain(data_train, label_train, "the", feature_names)}')
f'informa print(	ation gain at other keyword hillary is {compute_information_gain(data_train, label_train, "hillary", feature_names)}
	ation gain at other keyword trumps is {compute_information_gain(data_train, label_train, "trumps", feature_names)}')
informati informati	wangy\anaconda3\python.exe "C:/uoft-academic/2021 fall/CSC311/hw1/a1.py" on gain at root split donald is 0.052603317747226375 on gain at other keyword the is 0.047175200175596954 on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375 on gain at other keyword the is 0.047175200175596954
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459
informati informati informati	on gain at root split donald is 0.052603317747226375  on gain at other keyword the is 0.047175200175596954  on gain at other keyword hillary is 0.04127665563376459

Now that In (w) = $\frac{1}{2N} \frac{1}{12} (y \dot{\omega} - t^{(\dot{\omega})})^2 + \frac{1}{2} \sum_{i=1}^{N} \beta_i w_i^2$ and $y_i = \sum_{i=1}^{N} w_i y_i^2 + b - (i)$ substitute $y(\omega)$ in $J_{ni} f(\omega)$ with $(0)$ $J_{ni} f(\omega) = \frac{1}{2N} \sum_{i=1}^{N} ((\sum_{j=1}^{N} w_j x_j^{(\dot{\omega})} + b) - t^{(\dot{\omega})})^2 + \frac{1}{2} \sum_{j=1}^{N} \beta_j w_j^2$ break $J_{ni} f(\omega) = J_{ni} f(\omega) + J_{ni} f(\omega)$ ,  then, $J_{ni} J_{ni} f(\omega) = J_{ni} f(\omega) + J_{ni} f(\omega)$ calculate the partial derivative $W_i f(\omega)$ and $J_{ni} f(\omega) + J_{ni} f(\omega)$ $J_{ni} J_{ni} = \frac{1}{N} \sum_{i=1}^{N} ((\sum_{j=1}^{N} w_j x_j^{(\dot{\omega})} + b - t^{(\dot{\omega})}) \cdot J_{ni} f(\omega)$ $J_{ni} J_{ni} f(\omega) = J_{ni} f(\omega) \cdot J_{ni} f(\omega)$ $J_{ni} f(\omega) = J_{ni} f(\omega) \cdot J_{ni} f(\omega)$ Substitute $J_{ni} f(\omega) = J_{ni} f(\omega) \cdot J_{ni} f(\omega)$ $J_{ni} f(\omega) = J_{ni} f(\omega)$ $J_{ni} f(\omega) \cdot J_{ni$		
And that $\sum_{i=1}^{N} (y_i - t_i)^2 + \sum_{i=1}^{N} f_i w_j^2$ and $y_i = \sum_{i=1}^{N} (y_i - t_i)^2 + \sum_{i=1}^{N} f_i w_j^2$ Substitute $y_i y_i$ in $\int_{i=1}^{N} f_i(w)$ with $(0)$ $\int_{i=1}^{N} f_i(w) = \int_{i=1}^{N} f_i(w) + \int_{i=1}^{N} f_i(w)$ break $\int_{i=1}^{N} f_i(w) = \int_{i=1}^{N} f_i(w) + \int_{i=1}^{N} f_i(w)$ Calculate the partial derivative $W \cdot f_i(w)$ $\int_{i=1}^{N} \int_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_i X_i^{(i)} + b - t_i^{(i)} \right)^2 \right] = \int_{i=1}^{N} \int_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_i X_i^{(i)} + b - t_i^{(i)} \right) - f_i^{(i)} \right]$ $= \int_{i=1}^{N} \int_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_i X_i^{(i)} + b - t_i^{(i)} \right) - f_i^{(i)} \right]$ Substitute $f_i^{(i)}$ back: $= \int_{i=1}^{N} \int_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_i X_i^{(i)} + b - t_i^{(i)} \right) - f_i^{(i)} \right]$ Calculate the partial derivative $W \cdot f_i^{(i)}$ for $1 \le j \le N$ for $1 \le N$ for	Question 3	
Substitue $g(\omega)$ in $J_{reg}(\omega)$ with $(3)$ $J_{reg}(\omega) = \frac{1}{2}\omega \sum_{i=1}^{N} (\sum_{j=1}^{N} W_{i}x_{i}^{ij}+b) - t^{(i)})^{2} + \frac{1}{2}\sum_{j=1}^{N} B_{j}W_{j}^{2}$ break $J_{reg}(\omega) = J(\omega) + R(\omega)$ ,  then, $J_{reg}(\omega) = \frac{1}{2}\omega J(\omega) + \frac{1}{2}\omega R(\omega)$ Calculate the partial derivative $W_{reg}(\omega)$ for $I=j=D$ for $I(\omega)$ $J_{reg}(\omega) = \frac{1}{2}\omega \left( (\sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)}+b) - t^{(i)} \right)^{2}$ $= \frac{1}{N}\sum_{i=1}^{N} \left[ (\sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)}+b - t^{(i)}) \cdot \frac{1}{2}\omega \sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)} + \frac{1}{2}\omega \sum_{j=1}^{N} W_{j}^{(i)}X_{j}^{(i)} + \frac{1}{2}\omega \sum_{j=1}^{N}$	a)	know that N
Substitue $g(\omega)$ in $J_{reg}(\omega)$ with $(3)$ $J_{reg}(\omega) = \frac{1}{2}\omega \sum_{i=1}^{N} (\sum_{j=1}^{N} W_{i}x_{i}^{ij}+b) - t^{(i)})^{2} + \frac{1}{2}\sum_{j=1}^{N} B_{j}W_{j}^{2}$ break $J_{reg}(\omega) = J(\omega) + R(\omega)$ ,  then, $J_{reg}(\omega) = \frac{1}{2}\omega J(\omega) + \frac{1}{2}\omega R(\omega)$ Calculate the partial derivative $W_{reg}(\omega)$ for $I=j=D$ for $I(\omega)$ $J_{reg}(\omega) = \frac{1}{2}\omega \left( (\sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)}+b) - t^{(i)} \right)^{2}$ $= \frac{1}{N}\sum_{i=1}^{N} \left[ (\sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)}+b - t^{(i)}) \cdot \frac{1}{2}\omega \sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)} + \frac{1}{2}\omega \sum_{j=1}^{N} W_{j}^{(i)}X_{j}^{(i)} + \frac{1}{2}\omega \sum_{j=1}^{N}$		$\int_{eq}^{e}(w) = \frac{1}{2N} \sum_{i=1}^{e} \frac{(y^{i} - t^{(i)})^{2} + \frac{1}{2} \sum_{i=1}^{e} \beta_{i} W_{i}^{i}}{\sum_{i=1}^{e} \beta_{i} W_{i}^{i}}$
Substitue $g(\omega)$ in $J_{reg}(\omega)$ with $(3)$ $J_{reg}(\omega) = \frac{1}{2}\omega \sum_{i=1}^{N} (\sum_{j=1}^{N} W_{i}x_{i}^{ij}+b) - t^{(i)})^{2} + \frac{1}{2}\sum_{j=1}^{N} B_{j}W_{j}^{2}$ break $J_{reg}(\omega) = J(\omega) + R(\omega)$ ,  then, $J_{reg}(\omega) = \frac{1}{2}\omega J(\omega) + \frac{1}{2}\omega R(\omega)$ Calculate the partial derivative $W_{reg}(\omega)$ for $I=j=D$ for $I(\omega)$ $J_{reg}(\omega) = \frac{1}{2}\omega \left( (\sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)}+b) - t^{(i)} \right)^{2}$ $= \frac{1}{N}\sum_{i=1}^{N} \left[ (\sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)}+b - t^{(i)}) \cdot \frac{1}{2}\omega \sum_{j=1}^{N} W_{j}^{i}X_{j}^{(i)} + \frac{1}{2}\omega \sum_{j=1}^{N} W_{j}^{(i)}X_{j}^{(i)} + \frac{1}{2}\omega \sum_{j=1}^{N}$		and y= Ewixitb @
break $J_{eg}^{\mu}(w) = J(w) + R(w)$ , $then$ , $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w) + \frac{1}{dw}R(w)$ Calculate the partial derivative $W_{ef}^{\mu}(w) = \frac{1}{dw}J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w) + \frac{1}{dw}J(w) + \frac{1}{dw}J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w) + \frac{1}{dw}J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)J(w)J(w)J(w)J(w)J(w)J(w)$		Langui T Book and the
break $J_{eg}^{\mu}(w) = J(w) + R(w)$ , $then$ , $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w) + \frac{1}{dw}R(w)$ Calculate the partial derivative $W_{ef}^{\mu}(w) = \frac{1}{dw}J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w) + \frac{1}{dw}J(w) + \frac{1}{dw}J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w) + \frac{1}{dw}J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)J(w)J(w)J(w)$ $J_{eg}^{\mu}(w) = \frac{1}{dw}J(w)J(w)J(w)J(w)J(w)J(w)J(w)J(w)J(w)J(w)$		substitue y(a) in Jieg (W) with (3)
Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $J(w)$ $ \frac{1}{J} = \frac{1}{J} \underbrace{\frac{1}{J} \left( \left( \frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b} - t^{(i)} \right)^{i}}_{i = 1} \right) \left( \frac{J}{J} \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \right)^{i}}_{i = 1} \underbrace{\frac{J}{J} \underbrace{\frac{J}{J} \left( \left( \frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)} \right)^{i}}_{i = 1} \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \right) \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t$		Jrey (W) = 3 2 (13 W) (18) = (1 ) (2 ) (1 ) (1 )
Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $J(w)$ $ \frac{1}{J} = \frac{1}{J} \underbrace{\frac{1}{J} \left( \left( \frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b} - t^{(i)} \right)^{i}}_{i = 1} \right) \left( \frac{J}{J} \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \right)^{i}}_{i = 1} \underbrace{\frac{J}{J} \underbrace{\frac{J}{J} \left( \left( \frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)} \right)^{i}}_{i = 1} \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \right) \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t^{(i)}}_{i = 1} \cdot \underbrace{\frac{J}{J} \underbrace{W_{j}^{i} X_{j}^{(i)} + b}_{i = 1} - t$		break Treak(w) = T(w) + R(w).
Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $J(w)$ $J = \frac{1}{dw} \left( \frac{1}{2N} \sum_{i=1}^{N} \left( \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right)^{*} \right)^{*}$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot \frac{1}{dw} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + \frac{1}{dw} b - \frac{1}{dw} t^{(i)} \right)$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot \frac{1}{dw} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot X_{j}^{(i)} \right]$ Substitute $f^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot X_{j}^{(i)} \right]$ Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $A(w)$ $J(w) = \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ Then, $J(w) = \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$		then.
Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $J(w)$ $J = \frac{1}{dw} \left( \frac{1}{2N} \sum_{i=1}^{N} \left( \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right)^{*} \right)^{*}$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot \frac{1}{dw} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + \frac{1}{dw} b - \frac{1}{dw} t^{(i)} \right)$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot \frac{1}{dw} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot X_{j}^{(i)} \right]$ Substitute $f^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{2} \sum_{i=1}^{N} W_{i}^{*} X_{j}^{(i)} + b - t^{(i)} \right) \cdot X_{j}^{(i)} \right]$ Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $A(w)$ $J(w) = \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ Then, $J(w) = \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$ $J(w) = W_{j} - \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right) x_{j}^{*} + b_{j} w_{j}^{*}$		dy Jrey (w) = dus J(w) f dus R(w)
$ \frac{1}{\sqrt{N}} \int_{N} \frac{1}{\sqrt{N}} \left( \left( \sum_{j=1}^{N} W_{j}^{i} X_{j}^{(i)} + b \right) - t^{(i)} \right)^{2} $ $ = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \left( \sum_{j=1}^{N} W_{j}^{i} X_{j}^{(i)} + b - t^{(i)} \right) \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^{N} W_{i}^{i} X_{i}^{(i)} + \frac{1}{\sqrt{N}} \int_{N} W_{i}^{(i)} + \frac{1}{\sqrt{N}} \int_{N} W$		
$= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{j=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} + b - t \dot{\omega} \right) \left( \frac{1}{2} \int_{i=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} + \frac{1}{2} \int_{i=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} \right)$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{j=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} + b - t \dot{\omega} \right) \cdot \lambda_{j} \dot{\omega} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{N} - \frac{1}{N} \dot{\omega} \right) \cdot \lambda_{j}^{*} \dot{\omega} + b - t \dot{\omega} \right) \cdot \lambda_{j}^{*} \dot{\omega}$ Substitute $y^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y \dot{\omega} - t^{(i)} \right) \cdot \lambda_{j}^{*} \dot{\omega} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y \dot{\omega} - t^{(i)} \right) \cdot \lambda_{j}^{*} \dot{\omega} \right] = \beta_{j} W_{j}^{*}$ Then, $\frac{1}{N} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} + \beta_{j} w$ $\frac{1}{N} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} + \beta_{j} w$ $\frac{1}{N} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} - \alpha_{j} w$ $W_{j} = W_{j} - \alpha_{j} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} - \alpha_{j} w$ $W_{j} = W_{j} - \alpha_{j} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} - \alpha_{j} w$		Calculate the partial derivative Wirp W; for 15,50 for SCW)
$= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{j=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} + b - t \dot{\omega} \right) \left( \frac{1}{2} \int_{i=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} + \frac{1}{2} \int_{i=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} \right)$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{j=1}^{N} W_{j}^{*} X_{j}^{*} \dot{\omega} + b - t \dot{\omega} \right) \cdot \lambda_{j} \dot{\omega} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \frac{1}{N} - \frac{1}{N} \dot{\omega} \right) \cdot \lambda_{j}^{*} \dot{\omega} + b - t \dot{\omega} \right) \cdot \lambda_{j}^{*} \dot{\omega}$ Substitute $y^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y \dot{\omega} - t^{(i)} \right) \cdot \lambda_{j}^{*} \dot{\omega} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y \dot{\omega} - t^{(i)} \right) \cdot \lambda_{j}^{*} \dot{\omega} \right] = \beta_{j} W_{j}^{*}$ Then, $\frac{1}{N} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} + \beta_{j} w$ $\frac{1}{N} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} + \beta_{j} w$ $\frac{1}{N} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} - \alpha_{j} w$ $W_{j} = W_{j} - \alpha_{j} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} - \alpha_{j} w$ $W_{j} = W_{j} - \alpha_{j} \int_{i=1}^{N} \left( y \dot{\omega} - t^{(i)} \right) x_{j}^{*} \dot{\omega} - \alpha_{j} w$		d Tod (1 5 (1 2 Wox (1+L) -+ 6))
$= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_{j}^{i} \chi_{j}^{i} \dot{\omega} + b - t^{(i)} \right) \cdot \frac{d}{dw_{j}} \sum_{i=1}^{N} W_{i}^{i} \chi_{j}^{i} \dot{\omega} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_{j}^{i} \chi_{j}^{i} \dot{\omega} + b - t^{(i)} \right) \cdot \chi_{j}^{i} \dot{\omega} \right]$ Substitue $y^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y^{(i)} - t^{(i)} \right) \cdot \chi_{j}^{i} \dot{\omega} \right]$ Calculate the partial derivative $W \cdot r \cdot p$ $W_{j}$ for $1 = j \leq D$ for $R(W)$ $\frac{d}{dw_{j}} R(w) = \frac{d}{dw_{j}} \left( \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} + \beta_{j} w_{j} \right)$ Then. $\frac{d}{dw_{j}} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} + \beta_{j} w_{j}$ $\frac{d}{dw_{j}} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} - \alpha_{ij}^{N} \dot{\omega}$ $W_{j} = W_{j} - \alpha_{j}^{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} - \alpha_{ij}^{N} \dot{\omega}$ $W_{j} = W_{j} - \alpha_{j}^{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} - \alpha_{ij}^{N} \dot{\omega}$		This 3 - dw (21/2) (1/3) (1/3)
$= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_{j}^{i} \chi_{j}^{i} \dot{\omega} + b - t^{(i)} \right) \cdot \frac{d}{dw_{j}} \sum_{i=1}^{N} W_{i}^{i} \chi_{j}^{i} \dot{\omega} \right]$ $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \sum_{i=1}^{N} W_{j}^{i} \chi_{j}^{i} \dot{\omega} + b - t^{(i)} \right) \cdot \chi_{j}^{i} \dot{\omega} \right]$ Substitue $y^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y^{(i)} - t^{(i)} \right) \cdot \chi_{j}^{i} \dot{\omega} \right]$ Calculate the partial derivative $W \cdot r \cdot p$ $W_{j}$ for $1 = j \leq D$ for $R(W)$ $\frac{d}{dw_{j}} R(w) = \frac{d}{dw_{j}} \left( \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} + \beta_{j} w_{j} \right)$ Then. $\frac{d}{dw_{j}} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} + \beta_{j} w_{j}$ $\frac{d}{dw_{j}} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} - \alpha_{ij}^{N} \dot{\omega}$ $W_{j} = W_{j} - \alpha_{j}^{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} - \alpha_{ij}^{N} \dot{\omega}$ $W_{j} = W_{j} - \alpha_{j}^{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{i} \dot{\omega} - \alpha_{ij}^{N} \dot{\omega}$		= 1 2 ( 2 W x; a+h-+w)/ = 2 W x x a + d b - d + w)
Substitue $y^{(i)}$ back: $= \frac{1}{N} \sum_{i=1}^{N} \left[ \left( y^{(i)} + t^{(i)} \right) \cdot \chi_{j}^{(i)} \right]$ Calculate the partial derivative W.r.p Wj for $1 \le j \le D$ for $R(w)$ $\frac{d}{dw_{j}}R(w) = \frac{d}{dw_{j}} \left( \frac{1}{2} \sum_{j=1}^{N} \beta_{j} W_{j}^{2} \right) = \beta_{j} W_{j}$ then. $\frac{dw_{j}}{dw_{j}} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) \chi_{j}^{(i)} + \beta_{j} w_{j}$ $According the gradient descent$ $W_{j} = W_{j} - \alpha_{j} \int_{w_{j}}^{N} (y^{(j)} - t^{(j)}) \chi_{j}^{(i)} - \alpha_{j} w_{j}$ $W_{j} = W_{j} - \alpha_{j} \int_{w_{j}}^{N} (y^{(j)} - t^{(j)}) \chi_{j}^{(i)} - \alpha_{j} w_{j}$		-
Substitue $g^{(i)}$ back: $= \sqrt{\sum_{i=1}^{N}} \left[ y^{(i)} - t^{(i)} \right] \cdot x_{j}^{(i)}$ Calculate the partial derivative W·r.p Wj for $1 \le j \le D$ for $R(w)$ $= \frac{d}{dw_{j}} \left( x_{j}^{(i)} \right) = \frac{d}{dw$		= - = [( = W'X)" + b - tiu) - = = [ = W'X" ]
Substitue $g^{(i)}$ back: $= \sqrt{\sum_{i=1}^{N}} \left[ y^{(i)} - t^{(i)} \right] \cdot x_{j}^{(i)}$ Calculate the partial derivative W·r.p Wj for $1 \le j \le D$ for $R(w)$ $= \frac{d}{dw_{j}} \left( x_{j}^{(i)} \right) = \frac{d}{dw$		= - X [ (2 W!) X: W + L + W) . X: W
Calculate the partial derivative W.r.p W; for $1 \le j \le D$ for $R(W)$ $d_{Wj}R(W) = d_{Wj}\left(\frac{1}{2} \stackrel{?}{\ge} p_{j}W_{j}^{2}\right) = p_{j}W_{j}$ then. $d_{Wj}R(W) = \frac{1}{2} \stackrel{N}{\ge} (y^{ij} - t^{ij})x_{j}^{ij} + p_{j}W_{j}^{ij}$ $d_{Wj}R(W) = \frac{1}{2} \stackrel{N}{\ge} (y^{ij} - t^{ij})x_{j}^{ij} + p_{j}W_{j}^{ij}$ $d_{Wj}R(W) = \frac{1}{2} \stackrel{N}{\ge} (y^{ij} - t^{ij})x_{j}^{ij} + p_{j}W_{j}^{ij}$ $d_{Wj}R(W) = \frac{1}{2} \stackrel{N}{\ge} (y^{ij} - t^{ij})x_{j}^{ij} - \alpha p_{j}W_{j}^{ij}$ $d_{Wj}R(W) = \frac{1}{2} \stackrel{N}{\ge} (y^{ij} - t^{ij})x_{j}^{ij} - \alpha p_{j}W_{j}^{ij}$	Substitue UC	back:
Calculate the partial derivative W.r.p Wj for $1 \le j \le D$ for $R(w)$ $A_{wj}R(w) = A_{wj}\left(\frac{1}{2}\sum_{j=1}^{2}\beta_{j}W_{j}^{2}\right) = \beta_{j}W_{j}$ then. $A_{wj}R(w) = A_{wj}\left(\frac{1}{2}\sum_{j=1}^{2}\beta_{j}W_{j}^{2}\right) = \beta_{j}W_{j}$	or source (	$=\sqrt{2}\left[y\omega-t^{(i)}\right]\cdot x_{i}\omega$
$\frac{d}{dw_{j}}R(w) = \frac{d}{dw_{j}}\left(\frac{1}{2}\sum_{j=1}^{N}\beta_{j}W_{j}^{2}\right) = \beta_{j}W_{j}^{2}$ then. $\frac{d}{dw_{j}} = \frac{1}{ V }\sum_{i=1}^{N}(y^{ij}-t^{10})x_{j}^{ij} + \beta_{j}w_{j}$ $1x_{c} \text{ ording the gradient obescent}$ $W_{j} = W_{j} - \alpha_{j}V_{j}^{N}$ $W_{j} = W_{j} - \alpha_{j}V_{j}^{N}$ $W_{j} = W_{j} - \alpha_{j}V_{j}^{N}$		
$\frac{d}{dw_{j}}R(w) = \frac{d}{dw_{j}}\left(\frac{1}{2}\sum_{j=1}^{N}\beta_{j}W_{j}^{2}\right) = \beta_{j}W_{j}^{2}$ then. $\frac{d}{dw_{j}} = \frac{1}{ V }\sum_{i=1}^{N}(y^{ij}-t^{10})x_{j}^{ij} + \beta_{j}w_{j}$ $1x_{c} \text{ ording the gradient obescent}$ $W_{j} = W_{j} - \alpha_{j}V_{j}^{N}$ $W_{j} = W_{j} - \alpha_{j}V_{j}^{N}$ $W_{j} = W_{j} - \alpha_{j}V_{j}^{N}$		calculate the partial derivative W.r.y Wj for 1=j=D for R(W)
$\frac{\partial J_{reg}}{\partial w_{i}} = \frac{1}{1V} \sum_{i=1}^{N} (y^{ij} - t^{ij}) x_{j}^{ij} + \beta_{j} w_{j}$ $1 + c \text{ or dig the gradient obescent}$ $W_{i} = W_{i} - C + \frac{1}{2} \sum_{i=1}^{N} (y^{ij} - t^{ij}) x_{j}^{ij} - C + \frac{1}{2} W_{j}^{ij}$ $W_{i} = W_{i}^{i} - \frac{1}{1V} \sum_{i=1}^{N} (y^{ij} - t^{ij}) x_{j}^{ij} - C + \frac{1}{2} W_{j}^{ij}$		dn(1) d ( L Dr. 112 ) 2 1
$\frac{\partial J_{reg}}{\partial w_{i}} = \frac{1}{1V} \sum_{i=1}^{N} (y^{ij} - t^{ij}) x_{j}^{ij} + \beta_{j} w_{j}$ $1 + c \text{ or dig the gradient obescent}$ $W_{i} = W_{i} - C + \frac{1}{2} \sum_{i=1}^{N} (y^{ij} - t^{ij}) x_{j}^{ij} - C + \frac{1}{2} W_{j}^{ij}$ $W_{i} = W_{i}^{i} - \frac{1}{1V} \sum_{i=1}^{N} (y^{ij} - t^{ij}) x_{j}^{ij} - C + \frac{1}{2} W_{j}^{ij}$		Juik (W) = Jui (2 = P) Wi = ) = Pj Wi
He coording the gradient descent $W_i = W_i - \alpha \frac{1}{W_i} \frac{N}{N_i} (y^{(i)} - t^{(i)}) x_i^{(i)} - \alpha t_i w_i$		Then, Declared to the second that the second to the second
Wie Wi- Of HOUNG WING		The solid the condition of
Wie Wi- Of HOUNG WING		IN a Mi-N Ing
$W_{j} = W_{j} - W_{j} = W_{j} + W_{j} = \alpha U_{j} W_{j}$ $W_{j} = (1 - \alpha V_{j}) w_{j} - W_{j} = W_{j} + W_{j} = W_{j} + W_{j} = W_{j} = W_{j} + W_{j} = W_{j} $		
Wie (1-ap; )wj - TV Eq (yo-tw) xj (i)		Wie Wi- to E(yu-ti)xi w- all w
		Wie (1-ak; )wi - To Z(yo-tw) x; w
		· · ·

	Calculate of
	$\frac{dJ_{\text{rey}}^{\text{Rey}}}{db} = \frac{d}{db} \left( \frac{1}{2N} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} W_{j}^{i} / \chi_{j}^{i} \dot{\omega}^{i} + b - t^{(i)} \right)^{2} \right) + \frac{d}{db} \left( \frac{1}{2N} \sum_{j=1}^{N} B_{j} W_{j}^{i} \right)$
	$=\frac{d}{db}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}W_{j}'\chi_{j}^{i}(\lambda)+b-t^{(i)}\right)^{2}\right)$
	= 1/2 [( = W; 7; 4 + b - t10) ( db = W; 7; 4 + db - db (ta))]
Substitue Y (2) sha	$= \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} + b - t^{(i)} \right)$ $= \frac{1}{N} \sum_{i=1}^{N} \left( y^{(i)} - t^{(i)} \right)$
	According the gradient obscent $b \leftarrow b - \alpha \stackrel{\text{def}}{\Rightarrow} (y^{(i)} - t^{(i)})$
In conclusion:	Wi ≤ (1- ak; )w; - (1 = (y) - tw) x; iv
	b 6 b - # 2 (y () - t ())
Weight decay;	According to weight update form, we rescale the weights wij by (1-83;) wij on each gradient descent it refers to weight decay as it makes weights smaller
	it refers to weight decay as it makes weights smaller

b)	from parea), we know dirig = 1 = 1 (y (i) - t (i) x; (i) + B; W;
	Substitue y ij = Z Wj Xj i
	Jig = 1 2 ((2 Wj, xj, (0) - + 16) xj (0+ 1/2 Wj
	multiply both side by IV:
	$\sum_{i=1}^{N} \left( \left( \sum_{j=1}^{N} W_{j}^{i} \chi_{j}^{i,(i)} \right) - t^{(i)} \right) \chi_{j}^{(i)} + \beta_{j}^{i} W_{j}^{i} = \sqrt{\frac{d J r_{j}^{i}}{J w_{j}}}$
	Z [ Z Wj'xj' w xj w + Pj Wj ] - Z to xj w = N d Jreg )
Expressy PjW	as Zj'=   Wj'Bj'I; =1: Here I'=j'= is the indication function
	$\sum_{i=1}^{N} \left( \sum_{j=1}^{N} (X_{j}^{(i)} X_{j}^{(i)} + \beta_{j}^{*} I_{j=j}^{*}) w_{j}^{*} I_{j} - \sum_{i=1}^{N} t_{i}^{(i)} X_{j}^{*} (x_{j}^{*} - \lambda_{j}^{*}) d J r_{i}^{*} d w_{j}^{*} \right)$
	exchange order of is and is
	ELECX; Xj W+ B; J; j W; J- Et 10 Xj Ci) = / d Jreg)
	Accordy the form Z Ajj' Wj'-Cj = dry
	We con see that
	$A_{jj'} = \sqrt{\sum_{i=1}^{N} (x_{j'}^{ij} X_{j}^{ij} U + k_{j'} L_{j'} = j)}$
<i>f</i>	$Cj = V \stackrel{N}{=} t^{(i)}X_{j}^{(i)}$
	CJ = N Z T T Aj T

 $C = \frac{1}{N} \chi^{T} t$  (varified by matrix product) Similarly, first break  $A : J' = \frac{1}{N} \sum_{i=1}^{N} (\chi_{J}, (U_{\chi_{i}}, (U_{\chi_{i}}, U_{\chi_{i}}))$ + + ZNB" 512 First notice that is a (x; ("x; ") is the dot produce of j'and j'th col by aggregate all the sand i' which equal to XT-X to the second part. I Zin Ps' Is'=j = Bj' Is'=j iff i'=j by aggrevace all jund; in which aqual to diag (B) i.e a diagonal matrix on the jth position in the diagonal is Bj add the two parts. H= TXTX+dag(B) and C= TV XTE

Accordy the form  $= Aij'Wi'-Ci = \frac{dIreg}{dwj} = 0$ Neknow AW = C  $W = A^{-1}C$   $W = (X^{T}X + IV diag(B))^{-1}X^{T} + t$