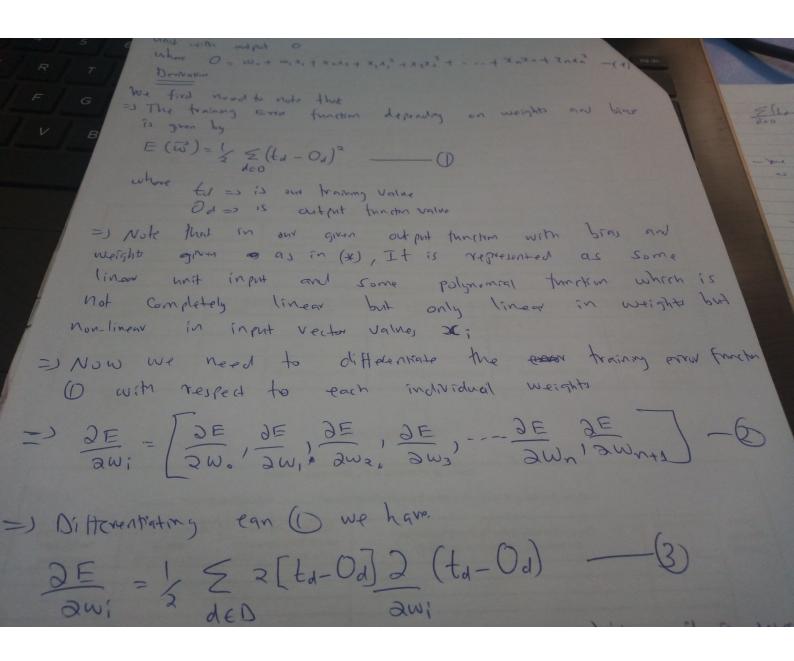
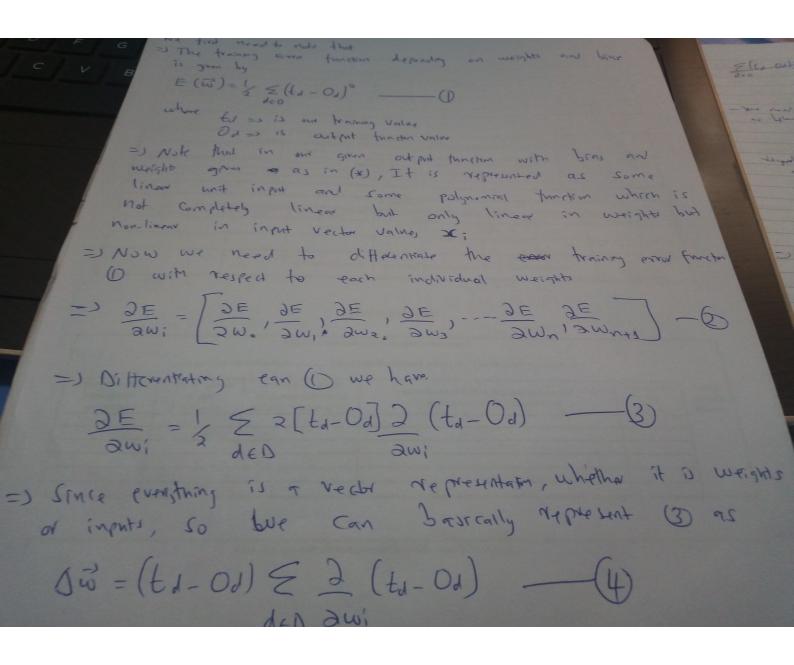


Derivation of gradient descent training vule for a single Unit with output 0 where 0 = wo + w, 2, + x2x2 + x, x, 2 + x2x2 + - - + xnxn + 2nxn Denvakin We first need to note that =) The training Error function depending on weights and brace is given by $E(\vec{w}) = \frac{1}{2} \leq (t_d - O_d)^2$ where to = > is our training value Od => 15 output fundon value = I Note that in our given output thritim with birds and weight given as as in (x), It is represented as some inow unit input and some polynomial tracken which

Derivation of gradient descent training vale for a single Unst with output o where 0 = wo + w, 2, + x2x2 + x, x, 2 + x2x2 + - - + xnxn + 2nxn We first need to note that = The training Error function depending on weights and $E(\vec{w}) = \left(\frac{1}{2} \leq (t_d - O_d)^2 - O\right)$ where Ed = 3 is out training value

Od = 3 is outfut funding value =) Note that in our given output tunition with bias and weight given as as in (x), It is represented as some linear unit input and some polynomial tracken which is Not completely linear but only linear in weight but Mon-linear in input vector value, of; I NOW WE need to differentiate the even training ever freeze O with respect to each individual weights SE = | SE 'SE 'SE '-- JE





Differentiating eqn () we have

Differentiating is a vector of the presentation which is away to the properties of the p

=) Question is what is Od is our case according the given out put truction (*) =) We need to write a general representation of and output function ors below $O_d = \overrightarrow{u} \times_a + \overrightarrow{u} \times_a^2 = O_a(\overrightarrow{x}) = \overrightarrow{u} \times_a + \overrightarrow{u} \times_a^2$ where wi is the weight vector that is general for all weights regardles of whether it is wo, we, wa, - so on Id is the input vector of each instance Substituting 60 into 60 we have = = [ta-0a] [ta-(\vec{v}\vec{x}_a + \vec{v}\vec{x}_a^3)] Now, differentiating ear 6 we have < [to-Od][0-Xi,d-Xi,d] JED

=) Question is what is Od is our case according the govern out put trunction (x) =) We need to write a general representation of and output function as below $O_d = \overrightarrow{u} \overrightarrow{X}_d + \overrightarrow{u} \overrightarrow{X}_d = O_a(\overrightarrow{x}) = \overrightarrow{u} \overrightarrow{X}_d + \overrightarrow{u} \overrightarrow{X}_d^2$ where we is the weight vector that is general for all weights regardles of whether it is wo, w, wa, - so Ky is the input vector of each instance J Substituting 6 into 6 we have = = [ta-0a] [ta-(wita+wita)] Now, differentiating ear 6 we have > \le \[\le \ta - O_d \] \[\le - \tilde \t Factory at regative in X; we have E1-00]-(X:1+X:1)

=) Questin a what is Od is -) We need to write a general representation output function as below $O_d = \vec{\omega} \vec{X}_d + \vec{\omega} \vec{X}_d = O_d(\vec{x}) = \vec{\omega} \vec{X}_d + \vec{\omega} \vec{X}_d$ where we is the weight vector that is general for all weights regardles of whether it is wo, we was - so on the input Vector of each instance = Substituting (5) into (we have = = [td - 0] [td - (\vec{v}\vec{x}_d + \vec{v}\vec{x}_d^3)] =) Now, differentiating ear 6 we have => \le [td - Od] [0 - \times i,d - \times i,d] => Factory out negative in x; we have $= \sum_{d \in D} \left[t_d - O_d \right] - \left(\overrightarrow{X}_{i,d} + \overrightarrow{X}_{i,d} \right)$ Ever trustom =) - $\sum [t_d - O_d][\vec{X}_{i,d} + \vec{X}_{i,d}] = \nabla E(\vec{u})$

out put truction (x) =) We need to write a general depresentation at output fraction as below $O_d = \vec{\omega} \vec{X}_d + \vec{\omega} \vec{X}_d = 0$ $O_a(\vec{x}) = \vec{\omega} \vec{X}_d + \vec{\omega} \vec{X}_d - 6$ where we is the weight vector that is general for all weights regardless of whether it is wo, we, wa, - so on the tre input vector at each instance = Substituting 6 into Q we have -6 = = [ta - 0a] = [ta - (\vec{v}\vec{x}_a + \vec{v}\vec{x}_a^3)] => Now, differentiating ear 6 we have => \le [to-Od][0-\tilde{X}_{i,d}-\tilde{X}_{i,d}] =) Factory at regative in x; we have $= \sum_{d \in D} \left[\frac{1}{X_{i,d}} + \overline{X_{i,d}} \right]$ Ever tructur =) - $\sum \left[\frac{1}{4} - Od \right] \left[\overrightarrow{X}_{i,d} + \overrightarrow{X}_{i,d} \right] = \nabla E(\vec{\alpha})$ - let's have a learning rate given by 2

=) D= - 2VE(2) => This can be represented in other notation DW: = JDE 3 Three training rule for gradient descent is w; ← w; + Aw; => Thosetore the modified weight equation is grown by w; = w, - 2 [td-0][X;d+X;p] dED

