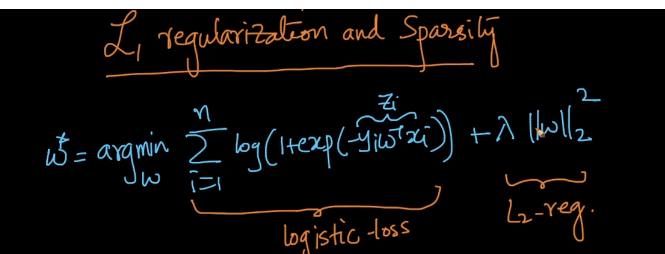
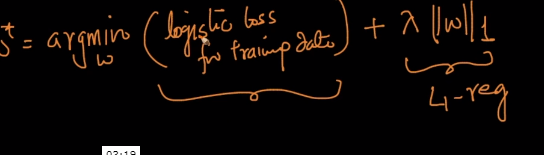
**L1 regularization and sparsity**

Just to recap, W\* is finding minimum w value by summation of logistic loss and L2 regularization. We use regularization to avoid z🡪∞ meaning avoiding Wi 🡪 ∞ or Wi 🡪 -∞  
  
Just like, above we have an L2 regularization, we have L1 regularization, this is alternative for L2 regularization. The use of it is, same as L2 regularization but have additional advantage called **sparsity**. Equation will change as below after adding L1 regularization.  
in L1 regularization, we use ||W||1  and we call it as 1 norm.  
1 norm of vector is part of minkowski distance and defined as below and we call it as absolute value of Wi for a ‘d’ dimension W  
  
**New definition** when L1 is used:  


**Sparsity :**Solution to LR is said to be sparse if man Wi’s are zero.  
When we are using L1 reg. in LR, all the unimportant or less important weights(Wi’s) will become Zero and this property is called sparse  
When we are using L2 regularization in LR, unimportant weight(Wi’s) becomes very small but not necessarily zero.

**So why does L1 reg. creates in W as compared to L2 ?**we can prove this statement using optimization.  
There is other alternative for L1 and L2, it’s called elastic-net. When we do this, we have two hyperparameters, Lamdba 1 and lambda2. We call this as **Elastic-net Logistic regression.**  
