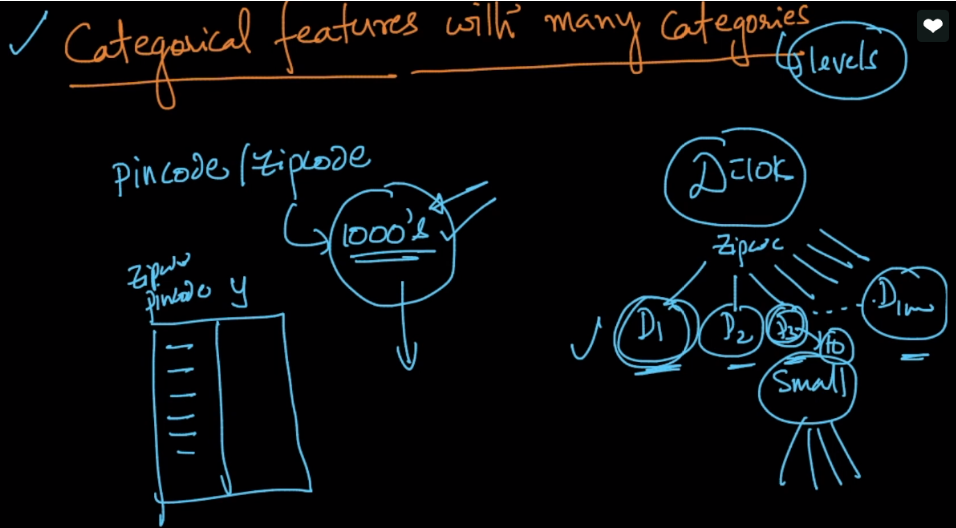
**How to deal when we have categorical features with many categories:**

Let’s say we have picode as a feature, as we know this is categorical feature, but there are many distinct pincodes. Therefore if we split according to pincode, there will be ‘n’ children for ‘n’ pincodes, and because of that at each child node there will be very less no. of points, or we can say it creates sparsity.



**So how to deal with this problem:**

So we’ll convert them into numerical values, but how can we convert categorical values into numerical values.

Let’s say we have output or dependent feature which can have two values 0 or 1.

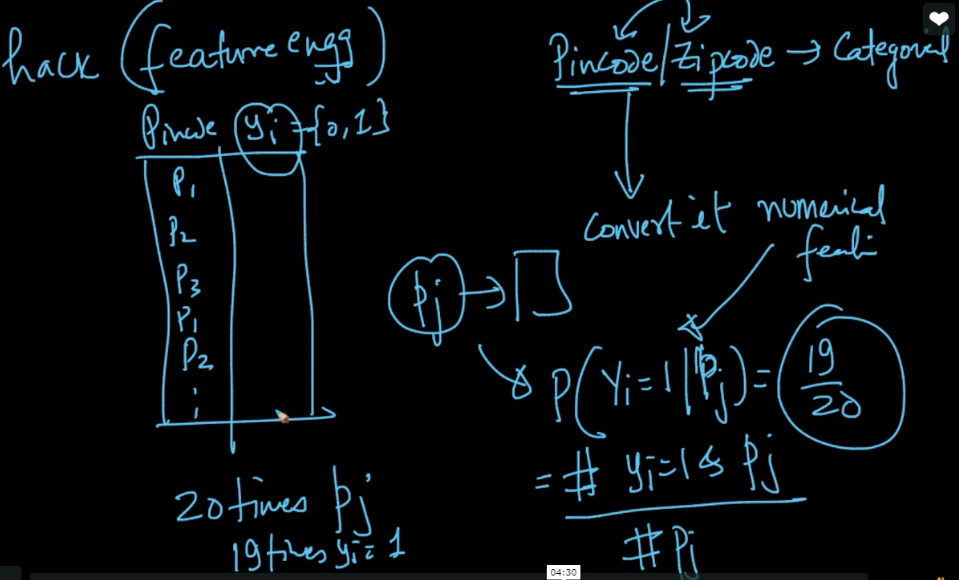
Now for each pincode p\_j we find probability P(y=1 | p\_j) ie probability of getting y = 1 given pincode is p\_j

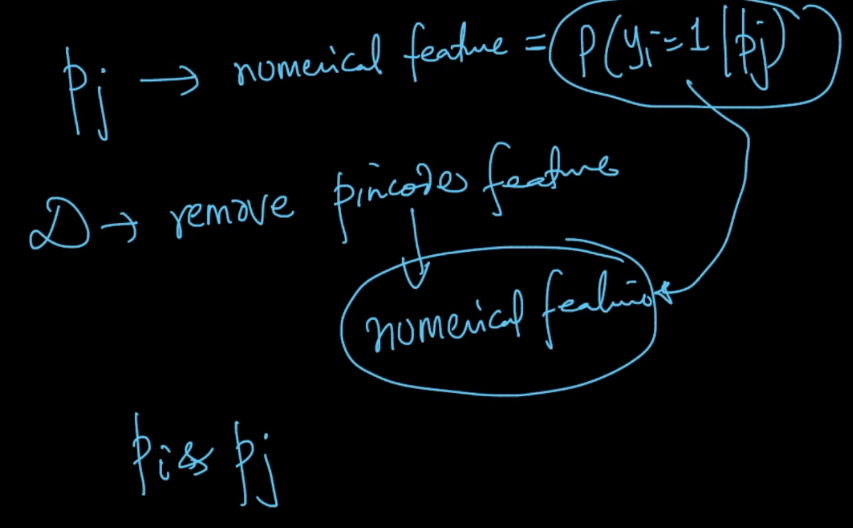
P(y=1 | p\_j) = (no. of data points where y = 1 and pincode = p\_j) / # where pincode = p\_j

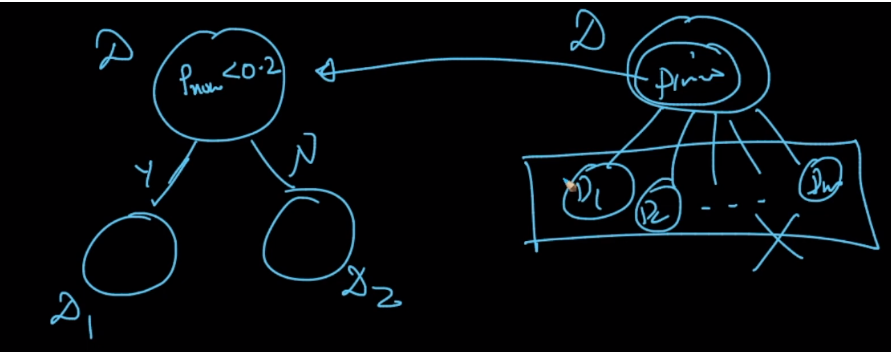
We’ll do this for all pincodes.

After finding this for all pincodes we’ll replace pincode column with this numerical values column.

Now we have numerical feature and now we can perform splitting as we do for numerical features

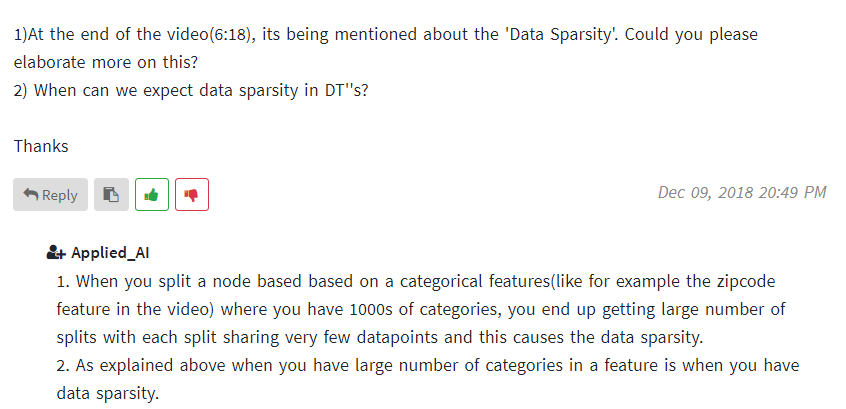




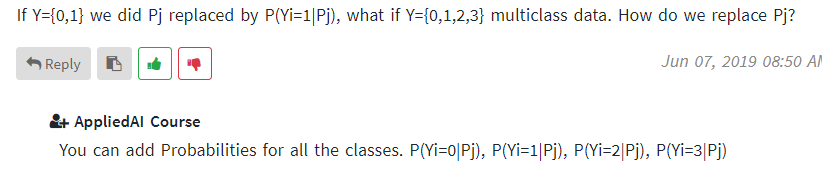


Comments:

1)



2) what if we have output feature containing values 0, 1, 2, 3



3)

