Unit 7\_7 update naïve Bayes model after adding one more training example

The original naïve Bayes model:

P(NO)=7/10, P(YES)=3/10

P(refund=yes|NO)=3/7

P(refund=no|NO)=4/7

P(maritalStatus=single|NO)=2/7

P(maritalStatus=divorced|NO)=1/7

P(maritalStatus=married|NO)=4/7

P(refund=yes|YES)=0/3

P(refund=no|YES)=3/3

P(maritalStatus=single|YES)=2/3

P(maritalStatus= divorced |YES)=1/3

P(maritalStatus= married |YES)=0

For taxable income,

if class=NO, sample mean=110, variance 2975;

if class=YES, sample mean=90, variance 25

Now update this model after adding the new example whose class is YES ((Refund=No, Married, Income=120K, Evade=Yes). The total number of examples changes to 11. So the class probabilities change:

P(NO)=7/11, P(YES)=3/11

The conditional probabilities for the YES class also need update, but those for the NO class remain the same:

P(refund=yes|YES)=0/4

P(refund=no|YES)=4/4

P(maritalStatus=single|YES)=2/4

P(maritalStatus= divorced |YES)=1/4

P(maritalStatus= married |YES)=1/4

For taxable income, for the YES class, the sample mean is changed to

sample mean=(95+85+90+120)/4=97.5  
sample variance=[(95-97.5)2+(85-97.5)2+(90-97.5)2+(120-97.5)2]=241.7

Model update cost: updating the naïve Bayes model will be faster than creating an original model. Only part of the naïve Bayes model will be updated when new training examples are coming, because only the prior category probabilities and the conditional probabilities relevant to the labels of the new examples are needed to be calculated. Furthermore, such computation takes one round and does not require iterations.