Corrections in Lecture Notes

I mentioned about the confusion in the lecture. The JHan's slide posted in the Lecture Notes uses the names of the sets differently than Kumar's book. However, both books meant the same ! It means that bootstrapped sample 63.2% of the original pool is used as a training set to derive a model and the left over 36.8% of the original pool is used as a test set to test the model in each iteration. JHan's slide posted on the web calls them the other way but it means the same because the original pool is actually called "original training set" data and each 63.2% of the data is used to test each model in each iteration, so it calls "test set". I think Kumar's book is better to name each set. Follow the Kumar's book if it confuses you.

Each sample has probability $(1 – 1/n)^n$ of being Selected should be

$1 - (1 – 1/n)^n$

Gini(N1)

$= 1 – (5/6)^2 – (2/6)^2$

$= 0.194$

Gini(N2)

$= 1 – (1/6)^2 – (4/6)^2$

$= 0.528$

=➜

Gini(N1)
$= 1 – (5/7)^2 – (2/7)^2$
$= 0.194$
Gini(N2)
$= 1 – (1/5)^2 – (4/5)^2$
$= 0.528$