

Agriculture data analytics

—

- We have built a bayes estimator that predicts the output problem based on the input crop details and weather information.
- There are 4000 advices in the input data that correspond to a crop problem.
- Of these advices, 10% instances are used as a test data set for the bayes estimator which is based on the probabilities calculated from the remaining data set.
- As of now, we have not calculated the cross validation accuracy.
- The probabilities that are used for our estimator are explained in the following slides.

First we identify the season and crop species from the input advice. Then we fix the season and the crop, and find the following probabilities:

- Probabilities of occurrence for all the problems i.e, $p(\text{problem})$ for given species of crop.
- Probability of weather parameter being at certain value, given a particular problem i.e, $p(\text{weather parameter}/\text{problem})$. We estimate this by finding the values of all the weather parameters corresponding to the dates of problem occurrence.
- Probability of the crop being at certain age, given a particular problem i.e, $p(\text{age of crop}/\text{problem})$.

Then, for each problem, we find this probability :

$$p(\text{problem}/[\text{temperature, humidity, age, other weather parameters}]) =$$

$$\frac{p([\text{temp, humidity, age, other parameters}]/\text{problem}) * p(\text{problem})}{p([\text{temp, humidity, age, other parameters}])}$$

$$= \frac{p(\text{temp}/\text{pb}) * p(\text{humidity}/\text{pb}) * p(\text{age}/\text{pb}) * p(\text{other}/\text{pb}) * p(\text{problem})}{p(\text{temp}) * p(\text{humidity}) * p(\text{age}) * p(\text{other parameters})}$$

(Assuming independence between different parameters)

- The problem with highest posterior probability will be the final output.
- The accuracy in the case of this particular test set is 70.549%
- Since the advice data is not uniform and there are only few major problems that constitute the most, we have not calculated the cross validation accuracy for this model yet.
- In the following screenshot, the problems on the right side are predicted outputs and those on the left side are the actual observations.

aphid aphid
thrips jassid
borer sucking pest
thrips thrips
mildew sucking pest
borer borer
thrips thrips
thrips thrips
mildew sucking pest
thrips thrips
jassid jassid
aphid sucking pest
borer borer
boll worm boll worm
thrips thrips
aphid aphid
aphid jassid
aphid aphid
thrips thrips
thrips jassid
helicoverpa helicoverpa
borer borer
sucking pest sucking pest
thrips thrips
borer borer
sucking pest sucking pest
sucking pest sucking pest
thrips thrips
jassid mites
thrips thrips
thrips thrips
thrips thrips
borer borer
aphid jassid
sucking pest borer
thrips thrips
jassid die back
thrips jassid
borer borer
Number of incorrect predictions:134
Number of input test vectors:455
Accuracy: 70.5494505495 %
dheeraj@dheeraj-Lenovo-G50-70:~/new/BTP/cur/new\$

Thank you