Week 5. Correlation rules

CONTENTS

- Popular correlation structures.
- Bayes approach to prediction and Naïve Bayes classifier.
- Assigning articles to categories: Naïve Bayes algorithm.
- Bag-of-words text model.
- Decision tree and splitting criteria.
- Metrics of accuracy.

Week 5. I: Correlation structures, 1

Typically, to analyze relations between different aspects**, all the features are divided in two parts:**

input features X (one aspect) and

target features U (another aspect).

Then a rule F is sought to establish a relation between the input and target features, most usefully like U=F(X). This would allow for **predicting U from X.**

Week 5. I: Correlation structures, 2

Rule U=F(X) can be used for prediction of U from X. Because of its practical importance, the problem has received huge attention by researchers.

Several interesting forms of rules have been explored. Among them:

Week 5. I: Correlation structures, 3

Among popular types of rule U=F(X):

(i) Econometric structural model

(ii) Hidden Markov chain

(iii) Bayes network

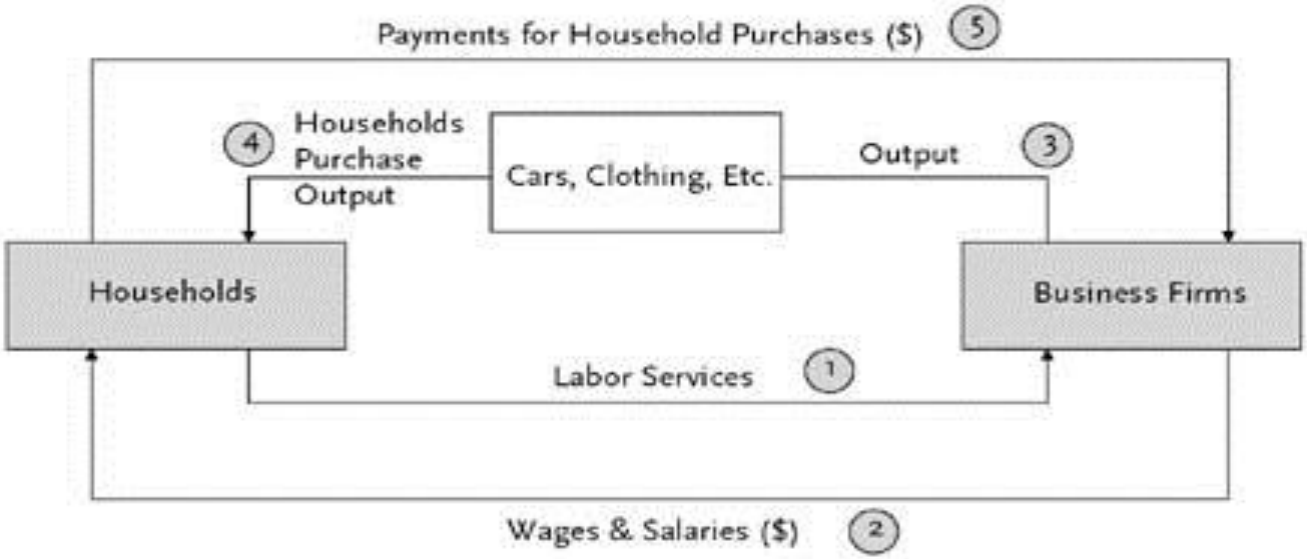
(iv) Neural network

(v) Decision tree

Week 5. I: Rule U=F(X) (i)

(i)Econometric structural model of dynamics:

features relate according to a predefined structure; some of them are exogenous (external), some endogenous (internal). Issues: (a) unmeasurable variables, (b) instability of processes, (c) non-linear, largely unknown, relations.



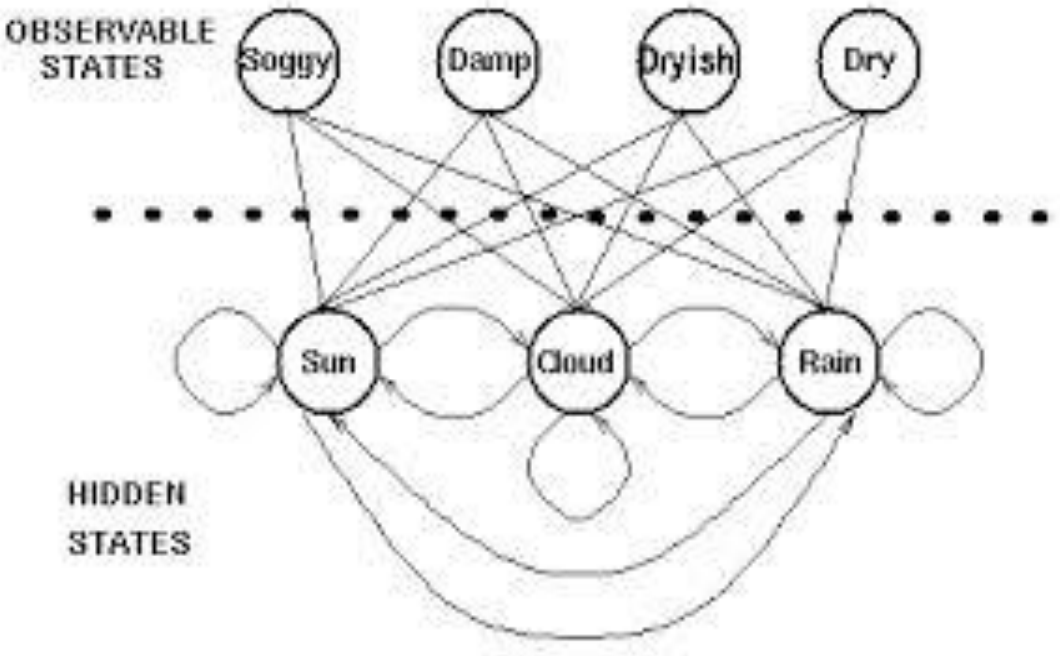
The diagram is from
http://www.econlib.org/library/Enc/ForecastingandEconometricModels.html#lfHendersonCEE2-062_figure_021

Week 5. I: Rule U=F(X) (ii)

(ii) Hidden Markov chain

Example: Using seaweed to predict the weather:

Using observable states to predict hidden states.



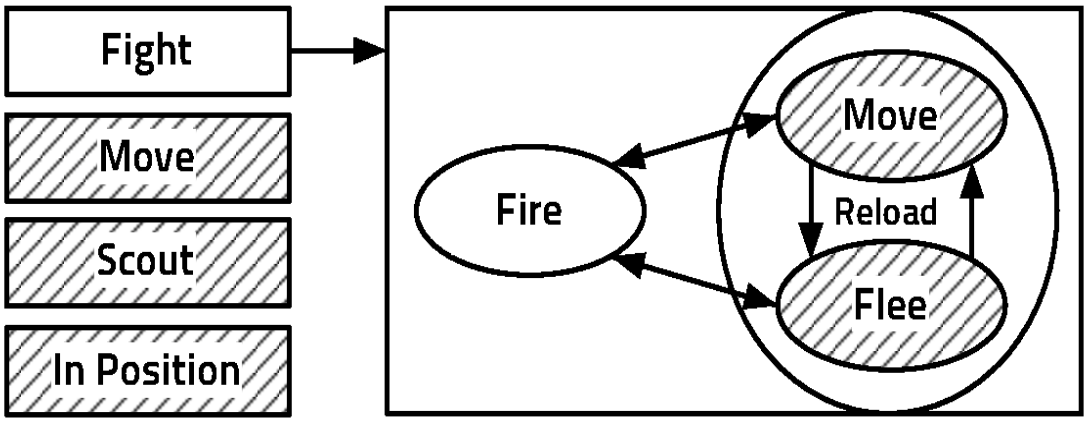
Hidden states form a simple probabilistic process, Markov chain, related to observable states by a confusion matrix. Issues: too complex to fit, too simple to model real world dynamics sometimes.

The diagram is from
http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/hidden_patterns/s2_pg1.html

Week 5. I: Rule U=F(X) (iii)

(iii) Bayes network

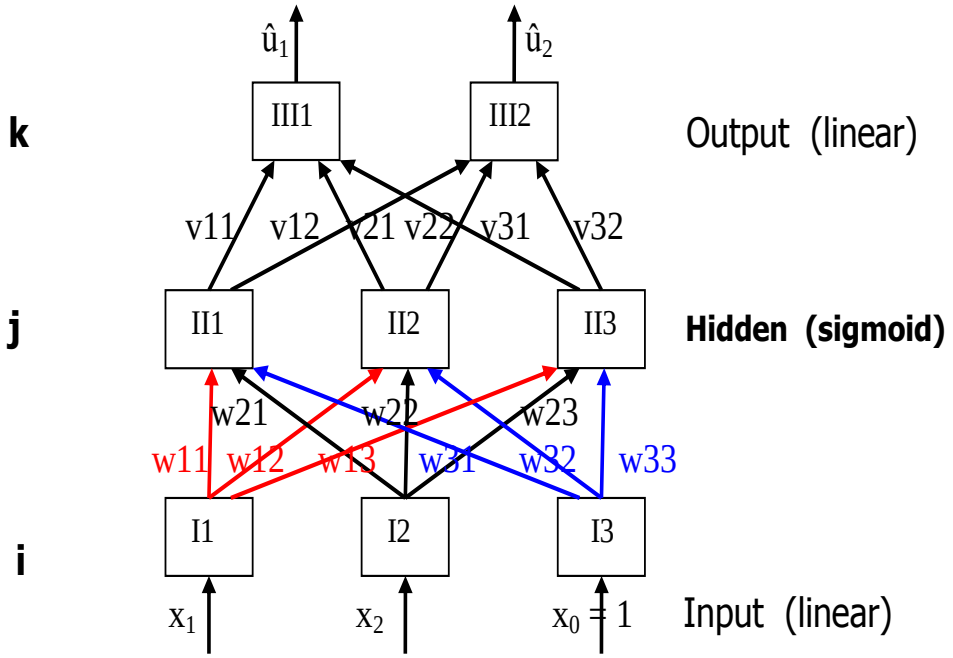
See: http://emotion.inrialpes.fr/people/synnaeve/phdthesis/phdthesis.html#x1-610003.2.2



Example: network of variables describing a class of computer games, with a dependence graph imposed on them. The outcome probabilities depend on those of others according to the graph structure**. Issue: finding a simple and adequate structure.**

Week 5. I: Rule U=F(X) (iv)

(iv) Neural network (diagram is from Mirkin 2011)



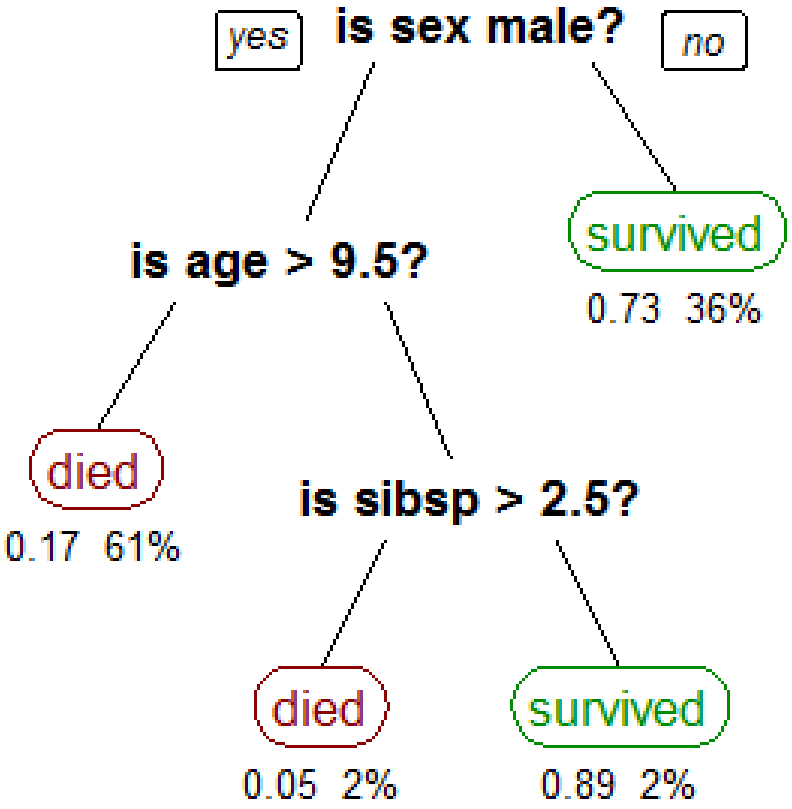
Feed-forward Neural Network for relating Iris sepal measures (x_1, x_2) with Iris petal measures (u_1, u_2). One hidden layer. **Issues: arbitrariness of the structure; lack of interpretation; inadequacy regarding real phenomena.**

Week 5. I: Rule U=F(X) (v)

(v) Decision tree

From http://en.wikipedia.org/wiki/Decision_tree_learning

A tree showing survival of passengers on the Titanic; "sibsp" is the number of spouses or siblings aboard. The figures under the leaves: the probability of survival and the percentage of observations.



Much interpretable, yet unreliable at weak correlation situations.

Week 5. I: Correlation structures, 4

I have presented five popular types of rule U=F(X):

(i) Econometric structural model

(ii) Hidden Markov chain

(iii) Bayes network

(iv) Neural network

(v) Decision tree

Two subjects, (iii) and (v), will be touched in more detail in the remainder of this week lecturing.

End of part 1 in Week 5

Week 5. II: Naïve Bayes classifier, 1

This relates to rule **(iii) Bayes network,** in Bayes part rather than network part ("network" with no links).

T. Bayes' (1702-1761) perspective at data analysis:

> **"There is a probabilistic distribution over patterns. Observed entities change that from a prior one to a posterior distribution."**

Week 5. II: Naïve Bayes classifier, 2

A database of 12×10 newspaper article-to-keyword items. Article labels: F for feminism, E for entertainment, and H for household.
Problem: Classify the last item x, of an unknown category.

Article	Keyword									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1
E1	2	0	1	2	2	0	0	1	0	0
E2	0	1	0	3	2	1	2	0	0	0
E3	1	0	2	0	1	1	0	3	1	1
E4	0	1	0	1	1	0	1	1	0	0
H1	0	0	2	0	1	2	0	0	2	0
H2	1	0	2	2	0	2	2	0	0	0
H3	0	0	1	1	2	1	1	0	2	0
H4	0	0	1	0	0	2	2	0	2	0
x	1	1	2	1	1	0	0	1	0	0

Week 5. II: Naïve Bayes classifier, 3

A database of 12×10 newspaper article-to-keyword items. Article labels: F for feminism, E for entertainment, and H for household.

Problem: Classify the last item x, of an unknown category.

Bayesian thinking:

Consider the prior situation according to the 12 item database: three classes, **F,E,** and **H**, probabilities of which are **p(F)=1/3, p(E)=1/3, p(H)=1/3**. Indeed of the 12 items, each class accounts for 4 of the, leading to 4/12 proportion.

Assume we can derive probabilities of **x** in each of the classes, **p(x|F), p(x|E), p(x|F)** from the database. Then the posterior probabilities of the classes will be proportional to the products (Bayes theorem): **P(F|x)=p(x|F)p(F), P(E|x)=p(x|E)p(E), P(H|x)=p(x|H)p(H).**

Therefore, the Bayesian solution:

Assign x to category with maximum posterior probability.

Week 5. II: Naïve Bayes classifier, 4

A database of 12×10 newspaper article-to-keyword items. Article labels: F for feminism, E for entertainment, and H for household.

Problem: **Classify the last item x, of an unknown category.**

Bayesian solution:

Assign x to category with maximum posterior probability, or that of maximum among

P(F|x)=p(x|F)p(F), P(E|x)=p(x|E)p(E), P(H|x)=p(x|H)p(H).

Caveat: How can we derive probabilities of **x** in each of the classes, **p(x|F), p(x|E), p(x|H)** from the database?

Week 5. II: Naïve Bayes classifier, 5

Caveat: How can we derive probabilities of **x** in each of the classes, **p(x|F), p(x|E), p(x|H)** from the database?

An universal advice: Be naïve!

The nature favors them. Similar principles: Occam razor, Maximum parsimony, Minimum length, Maximum likelihood. Of course each of these is utterly wrong, yet conclusions derived from them are quite reasonable.

Week 5. II: Naïve Bayes classifier, 6

Caveat: How can we derive probabilities of query **x** in each of the classes, **p(x|F), p(x|E), p(x|H)** from the database?

Naïve Bayes principle: assume that the features are independent within each class F, E, H:

1. drink
2. equal
3. fuel
4. play
5. popular
6. price
7. relief
8. talent
9. tax
10. woman

Then, given probabilities $f_1, f_2,\ldots,f_{10}$ of the keywords within each class, a product of them according to the query would do!

End of part 2 in Week 5

Week 5. III: Naïve Bayes classifier rule, 1

Naïve Bayes principle: assume the keywords are independent within each class of articles.

Given:

- Probabilities of ***1***-st, ***2***-nd ,…, ***m***-th keywords f_{k1}, f_{k2},…, f_{km} at ***k***-th class
- Query $x=(x_1, x_2,\ldots,x_m)$ where x_t is the number of occurrences of ***t***-th keyword,

Probability ***P(x|k)*** of ***x*** at **k**-th class is

$$P(x|k) = f_{k1}^{x_1} f_{k2}^{x_2} \cdots f_{km}^{x_m}$$

because of the independence assumption

Week 5. III: Naïve Bayes classifier rule, 2

Naïve Bayes algorithm.

Given a database of classified articles with ***m*** keywords and query ***x=(x$_1$, x$_2$,…,x$_m$)*** where ***x$_t$*** is the number of occurrences of ***t***-th keyword

1. Compute prior class probabilities ***P(k)***, ***k=1,2,…,K***.
2. Compute probabilities of ***1***-st, ***2***-nd ,…, ***m***-th keywords ***f$_{k1}$, f$_{k2}$,…, f$_{km}$*** at each ***k***-th class (***k=1,.., K***).
3. Compute probability ***P(x|k)*** of ***x*** at class ***k***

$$P(x|k) = f_{k1}^{x_1} f_{k2}^{x_2} \cdots f_{km}^{x_m} \qquad (*)$$

4. Compute products ***P(k|x)=P(k)P(x|k)*** and assign ***x*** to that class **k** for which ***P(k|x)*** is maximum.

Week 5. III: Naïve Bayes classifier rule, 3

Naïve Bayes algorithm: a PITFALL

Given a database of classified articles with ***m*** keywords and query ***x=(x$_1$, x$_2$,…,x$_m$)*** where ***x$_t$*** is the number of occurrences of ***t***-th keyword

3. Compute probability ***P(x|k)*** of ***x*** at class ***k***

$$P(x|k) = f_{k1}^{x_1} f_{k2}^{x_2} \cdots f_{km}^{x_m} \quad (*)$$

Pitfall: The probabilities f_{kt} are small in applications, of the order of a thousandth or millionth.

The product (*) is **very** small then. Say, a product of a dozen of reals of the order of 0.001 each, will come out as a real of the order of unity divided by $\mathbf{10^{36}}$. This is a digital **zero** for all practical purposes.

Way out: Use logarithms instead.

Week 5. III: Naïve Bayes classifier rule, 4

Computationally sound version of Naïve Bayes algorithm.

1. Compute prior class probabilities ***P(k)***, ***k=1,2,…,K***.
2. Compute probabilities of ***1***-st, ***2***-nd ,…, ***m***-th keywords f_{k1}, f_{k2},…, f_{km} at each ***k***-th class (***k=1,.., K***).
3. Compute logarithm *of* probability ***P(x|k)*** of ***x*** at **k**-th class

$$LP(x|k) = x_1 log(f_{k1}) + x_2 log(f_{k2}) + \cdots + x_m log(f_{km})$$

4. Compute sums

$$LP(k|x) = log(P(k)) + LP(x|k)$$

and assign ***x*** to that class **k** for which ***LP(k|x)*** is maximum.

End of Part 3 in Week 5

Week 5. IV: Probabilities of keywords,1

A crucial part of Naïve Bayes algorithm:

2. Computing probabilities of keywords within classes.

Article	Keyword									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1
E1	2	0	1	2	2	0	0	1	0	0
E2	0	1	0	3	2	1	2	0	0	0
E3	1	0	2	0	1	1	0	3	1	1
E4	0	1	0	1	1	0	1	1	0	0
H1	0	0	2	0	1	2	0	0	2	0
H2	1	0	2	2	0	2	2	0	0	0
H3	0	0	1	1	2	1	1	0	2	0
H4	0	0	1	0	0	2	2	0	2	0

Week 5. IV: Probabilities of keywords, 2

A crucial part of Naïve Bayes algorithm:

2. Computing probabilities of keywords within classes.

Take a look, say, at class F and see what is going on.

Article	Keyword									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1

First issue: Zero probability for keywords “fuel”, “relief”, and “tax” because of no occurrences. Not good because it would never assign an article x from the class to it if the article contains any of these keywords.

Week 5. IV: Probabilities of keywords, 3

A crucial part of Naïve Bayes algorithm:

2. Computing probabilities of keywords within classes.

Take a look, say, at class F and see what is going on.

Article	Keyword									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1

Second issue: What probability should one assign to “woman”? A unity? And what is about “popular” or “equal”? How one should take into account the multiple occurrences?

Week 5. IV: Probabilities of keywords, 4

A crucial part of Naïve Bayes algorithm:

2. Computing probabilities of keywords within classes.

Article	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1

"Bag of words model"

Put in a "bag" all the keywords from the upper line of the table (the 10 of them). Add all the occurrences of all keywords in the category (3+5+0+2+2+3+0+5+ 0+7=27), 37 altogether. The probability of a keyword, say, "equal" **is its total occurrence number**, 5, **plus 1**, **related to the bag size**, 37: 6/37=0.1622.

Week 5. IV: Probabilities of keywords, 5

A crucial part of Naïve Bayes algorithm:

2. Computing probabilities of keywords within classes with Bag of words model.

Class	Probabilities of keywords within classes									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F	0.108	0.162	0.027	0.081	0.081	0.108	0.027	0.162	0.027	0.216
E	0.095	0.071	0.095	0.167	0.167	0.071	0.095	0.143	0.048	0.048
H	0.049	0.024	0.171	0.098	0.098	0.195	0.146	0.024	0.171	0.024

For example, compute $f_{drink,E}=(3+1)/(32+10)=4/42=0.0952$.
Here 3 is the total number of occurrences of keyword "drink" in class E, and 32, the total number of occurrences of all the ten keywords in E, so that 42 is the size of bag of words for class E.

Week 5. IV: Probabilities of keywords, 6

A crucial part of Naïve Bayes algorithm:

Class	Within class probabilities of keywords									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F	0.108	0.162	0.027	0.081	0.081	0.108	0.027	0.162	0.027	0.216
E	0.095	0.071	0.095	0.167	0.167	0.071	0.095	0.143	0.048	0.048
H	0.049	0.024	0.171	0.098	0.098	0.195	0.146	0.024	0.171	0.024

Take logarithms (natural here) of the probabilities above expressed in hundredth (to make all the logarithms positive):

Class	Logarithms of within class probabilities of keywords multiplied by 100									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F	2.380	2.786	0.994	2.093	2.093	2.380	0.994	2.786	0.994	3.074
E	2.254	1.966	2.254	2.813	2.813	1.966	2.254	2.659	1.561	1.561
H	1.585	0.892	2.838	2.278	2.278	2.971	2.683	0.892	2.838	0.892

Week 5. IV: Naïve Bayes - Final decision,1

Class	Logarithms of within class probabilities of keywords multiplied by 100									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F	2.380	2.786	0.994	2.093	2.093	2.380	0.994	2.786	0.994	3.074
E	2.254	1.966	2.254	2.813	2.813	1.966	2.254	2.659	1.561	1.561
H	1.585	0.892	2.838	2.278	2.278	2.971	2.683	0.892	2.838	0.892
X	1	1	2	1	1	0	0	1	0	0

1. Take logarithms of 100/3 (class probabilities being equal here): C=log(100/3)=3.5066.

2. Take vector of query *x* (bottom line in Table above) and compute inner product of it and each of the lines in Table:

3. Add respective results of 1 and 2.

4. Assign x to class of maximum of the found values.

Week 5. IV: Naïve Bayes - Final decision,2

Class	Logarithms of within class probabilities of keywords multiplied by 100									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F	2.380	2.786	0.994	2.093	2.093	2.380	0.994	2.786	0.994	3.074
E	2.254	1.966	2.254	2.813	2.813	1.966	2.254	2.659	1.561	1.561
H	1.585	0.892	2.838	2.278	2.278	2.971	2.683	0.892	2.838	0.892
X	1	1	2	1	1	0	0	1	0	0

2. **Take vector of query *x* (bottom line in Table above) and compute inner product of it and each of the lines in Table:**

AF=1*2.380+1*2.786+2*0.994+1*2.093+1*2.093+0*2.380+0*0.994+1*2.786+0*0.994+0*3.074=14.127

AE=1*2.254+1*1.966+2*2.254+1*2.813+1*2.813+0*1.966+0*2.254+1*2.659+0*1.561+ 0*1.561 = 17.014

AH=1*1.585+1*0.892+2*2.838+1*2.278+1*2.278+0*2.971+0*2.683+1*0.892+0*2.838+0*0.892 = 13.600

Week 5. IV: Naïve Bayes - Final decision,3

Class	Logarithms of within class probabilities of keywords multiplied by 100									
	drink	equal	fuel	play	popular	price	relief	talent	tax	woman
F	2.380	2.786	0.994	2.093	2.093	2.380	0.994	2.786	0.994	3.074
E	2.254	1.966	2.254	2.813	2.813	1.966	2.254	2.659	1.561	1.561
H	1.585	0.892	2.838	2.278	2.278	2.971	2.683	0.892	2.838	0.892
X	1	1	2	1	1	0	0	1	0	0

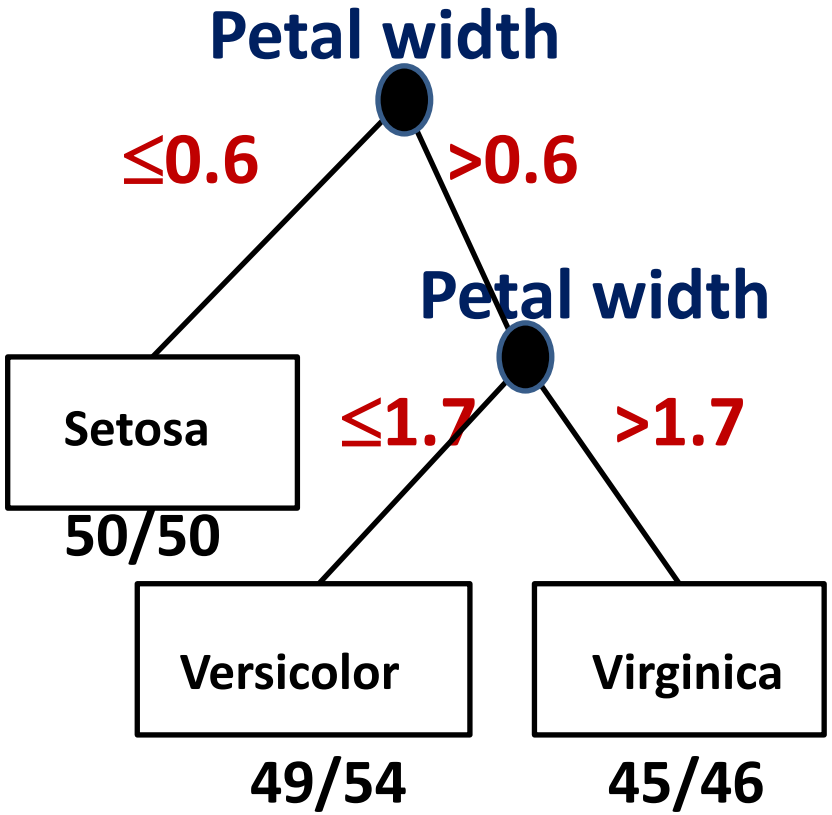
3. Add respective results of 1 and 2:

AF+C=17.633; **AE+C=20.520**; AH+C=17.105

4. Assign **x** to class of maximum value, **E**, in this case. (Indeed, **x** is for an article from **E**)

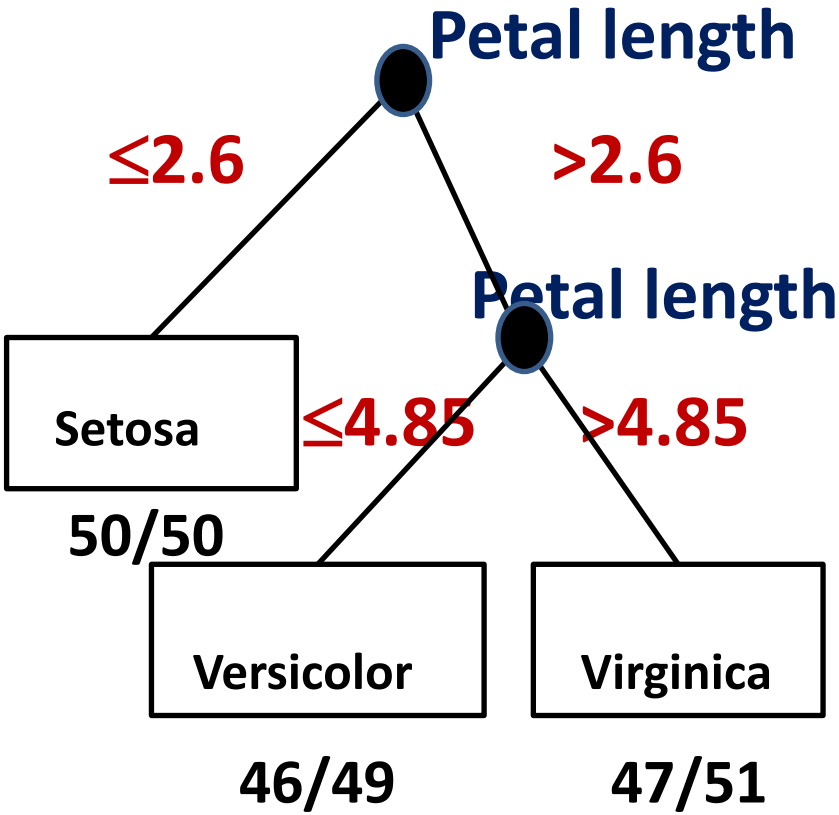
End of Part 4 in Week 5

Week 5. Correlation rules Part 5
Classification trees over Iris



6 errors

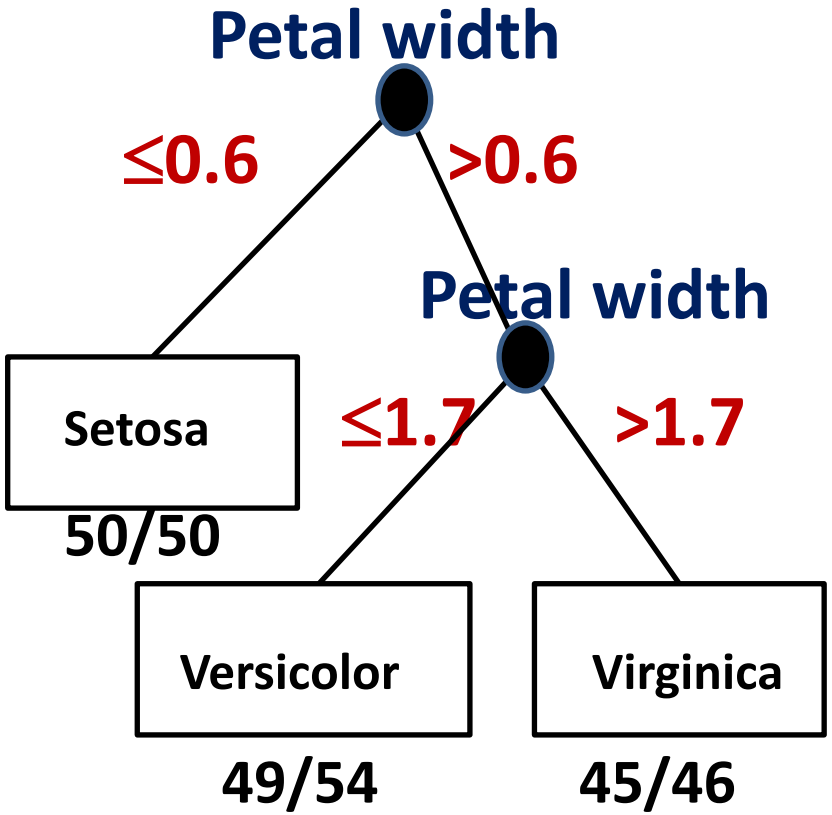
From Mirkin (2011)



7 errors

From http://www.ibm.com/developerworks/library/ba-predictive-analytics2

Week 5. Part 5. Classification tree,1



6 errors

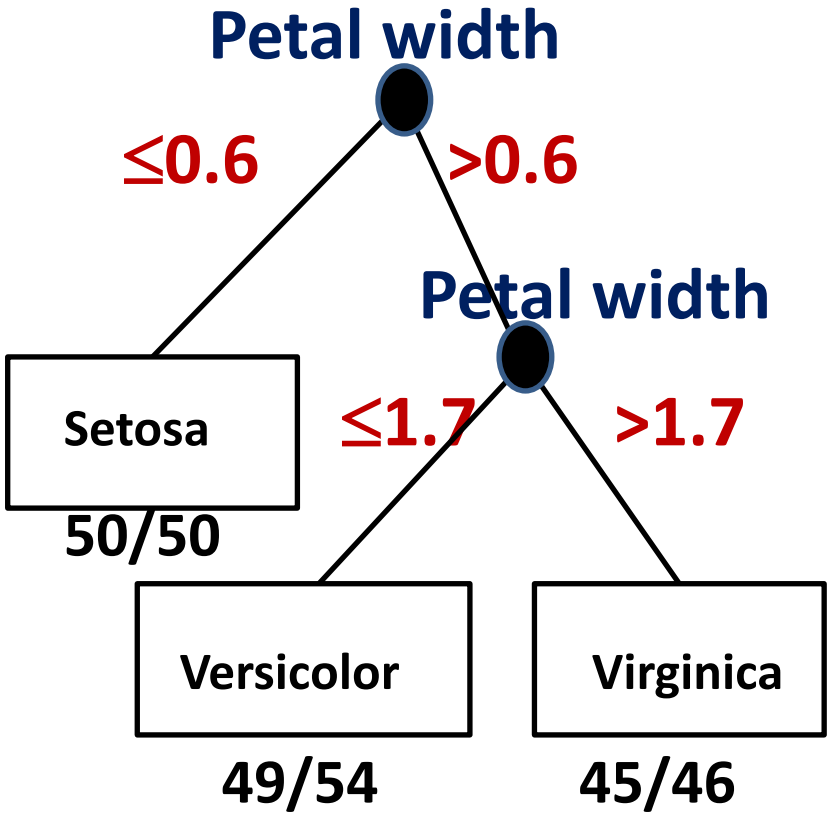
From Mirkin (2011)

Classification tree building over a training set, given a target partition *H*.

Goal: to build a partition *G* maximally similar to *H* by sequentially splitting a cluster over a feature.

Start: *G* consisting of one cluster, the dataset.

Week 5. Part 5. Classification tree,2



From Mirkin (2011)

Classification tree building over a training set, given a target partition *H*.

A split is chosen as the best of the splits of all clusters available over all feature values available.

Week 5. Part 5. Classification tree,3

A split is chosen as the best of the splits of all clusters available over all feature values available.

Split scoring function evaluates the similarity between the target partition *H* and partition *G* being built. Among most popular scoring functions are Pearson chi-squared (this is used in popular statistics SPSS package, for example) , Information gain, Category utility function.

Week 5. Part 5. Split scoring, 1

A unifying framework for split scoring functions:

1. **Take contingency table between target *H* and *G* being built *p(Hk∩Gl)* (co-occurrence frequency).**

2. Define a reasonable function *f(p(Hk∩Gl))* evaluating the extent of correlation between clusters *Hk* and *Gl*.

3. Score the similarity between G and H as the sum of all *f(p(Hk∩Gl))* weighted by their frequencies *p(Hk∩Gl).*

The mentioned Pearson chi-squared, Information gain, Category utility function are of this type.

Week 5. Part 5. Split scoring, 2

A unifying framework for split scoring functions:

Take the co-occurrence frequencies *p(Hk∩Gl)* between target *H* partition and *G* partition being built and define:

$$q(Hk|Gl) = \frac{p(Hk \cap Gl)}{p(Hk)p(Gl)} - 1$$ **Quetelet index** **(i)**

$$ig(Hk, Gl) = -log(\frac{p(Hk \cap Gl)}{p(Hk)p(Gl)})$$ **Mutual entropy** **(ii)**

$$a(Hk|Gl) = \frac{p(Hk \cap Gl)}{p(Gl)} - p(Hk)$$ **Quetelet difference** **(iii)**

Mathematical facts: Pearson chi-squared is the averaged *q* (i), Information gain is the averaged *ig* (ii), Category utility function is the averaged *a* (iii).

Week 5. Part 5. Split scoring, 3

Another unifying theme for split scoring functions:

$$q(Hk|Gl) = \frac{p(Hk \cap Gl)}{p(Hk)p(Gl)} - 1$$ **Quetelet index** **(i)**

$$a(Hk|Gl) = \frac{p(Hk \cap Gl)}{p(Gl)} - p(Hk)$$ **Quetelet difference** **(iii)**

Both Pearson chi-squared, the averaged *q* (i), and Category utility function, the averaged *a* (iii), are the contributions of partitions *H* and *G* to data scatter. The former, for the case when corresponding dummy features are normalized by $\sqrt{p_k}$ where p_k is the relative frequency of the target *Hk* category; the latter, not normalized.

This makes it **reasonable to apply Pearson chi-squared to the cases at which the more frequent categories** are expected to be less important, whereas the **Category utility function** applies when the **importance of categories has nothing to do with their frequency,**

Week 5. Part 5. An application: lift chart

In marketing research: lift is the relative response rate of a segment of population to an advert.

Cluster share, %	10	40	25	25
Response rate, %	30	10	4	0

Consider a sample divided in four clusters with response rates as in Table.

Baseline response rate is the mean:

0.1*30+0.4*10+0.25*4+0.25*0=8%

The lift of the first cluster is then 30/8=3.75, and that of the first two clusters is 14/8=1.75 because the response rate in their union is 14%.

End of part 5 in Week 5

Week 5. Part 6. Metrics of Accuracy, 1

Consider a classifier like that presented in Table, a lung screening device to diagnose lung cancer.

		True lung cancer		Total
		Yes	No	
Device's diagnosis	Yes	94	7	101
	No	1	98	99
Total		95	105	200

There 1+7=8 errors out of 200. Therefore, the *accuracy* is 192/200=96%, and *error* is 4%. These are too general and can be misleading sometimes.

Week 5. Part 6. Metrics of Accuracy, 2

Lung screening device errors and accuracy.

		True lung cancer		Total
		Yes	No	
Device's Diagnosis	Yes	94	7	101
	No	1	98	99
Total		95	105	200

1) There are two kinds of error: 7 False Positives and 1 False Negative, which may be of different cost. A FP will be identified in additional tests, whereas a FN may cause a lot trouble because of a late diagnosis.

Week 5. Part 6. Metrics of Accuracy, 3

Lung screening device errors and accuracy.

		True lung cancer		Total
		Yes	No	
Device's Diagnosis	Yes	94	7	101
	No	1	98	99
Total		95	105	200

2) There can be difference in correctly identified cases, True Positives and True Negatives, when a less balanced sample is screened.

Week 5. Part 6. Metrics of Accuracy, 4

Lung screening device errors and accuracy at a crowd at large.

		True lung cancer		Total
		Yes	No	
Device's Diagnosis	Yes	2	2	4
	No	1	195	196
Total		3	197	200

2) There can be difference in correctly identified cases, True Positives and True Negatives, because of an imbalanced sample. The *accuracy* here is high, 197/200=98.5%. But 1/3 of cancer sufferers is misdiagnosed, and ½ of "Yes" diagnoses are wrong.

Week 5. Part 6. Metrics of Accuracy, 5

Classifier errors and accuracy in general.

		True lung cancer		Total
		Yes	No	
Device's Diagnosis	Yes	TP	FP	TP+FP
	No	FN	TN	FP+TN
Total		TP+FN	FN+TN	All

Accuracy: $\frac{TP+TN}{All}$

Precision: $\frac{TP}{TP+FP}$ **(out of the classifier)**

Recall: $\frac{TP}{TP+FN}$ **(out of the classified)**

Week 5. Part 6. Metrics of Accuracy, 6

Lung screening device characteristics at a crowd at large.

		True lung cancer		Total
		Yes	No	
Device's Diagnosis	Yes	2	2	4
	No	1	195	196
Total		3	197	200

Accuracy=98.5%

Precision: $\frac{TP}{TP+FP} = \frac{2}{4} = 50\%$ **(out of the classifier)**

Recall: $\frac{TP}{TP+FN} = \frac{2}{3} = 67\%$ **(out of the classified)**

An average characteristic?

Week 5. Part 6. Metrics of Accuracy, 7

Averaging precision and recall

		True lung cancer		Total
		Yes	No	
Device's Diagnosis	Yes	2	2	4
	No	1	195	196
Total		3	197	200

Accuracy=98.5%

Precision: $\frac{TP}{TP+FP} = \frac{2}{4}$ = 50% (out of the classified)

Recall: $\frac{TP}{TP+FN} = \frac{2}{3} = 67\%$ (out of the classified)

F-measure= $\frac{2}{1/Precision+1/Recall} = \frac{2}{4/2+3/2} = \frac{4}{7} = 57\%$,

That is Harmonic mean

End of Part 6 in Week 5

Week 5. Correlation rules

Summary of subjects covered

1. **Popular correlation structures:** Five of them mentioned - Econometric structural model, Hidden Markov chain, Bayes network, Neural network, Decision tree

2. **Bayes approach to prediction and Naïve Bayes classifier:** Posterior probability can be a powerful tool.

3. **Assigning articles to categories:** Naïve Bayes algorithm employing a wrong assumption that the features are independent; using logarithms in it to avoid "infinitely" small numbers.

4. **Bag-of-words text model:** a simple device to estimate individual keyword probabilities.

Week 5. Correlation rules

Summary of subjects covered

5. **Decision tree and splitting criteria**: uncommon properties of popular splitting criteria including an assumption underlying the chi-squared that the feature weight is inversely related to its frequency.

6. **Metrics of accuracy**: an innate feature – there are two different types of error reflected in the measures of recall and precision.