

COMP30120 Tutorial

Decision Trees and Naïve Bayes Classifiers

Derek Greene

School of Computer Science and Informatics
Autumn 2015



Reminder: Entropy & Information Gain

- **Entropy** provides a measure of impurity - how uncertain we are about the decision for a given set of examples.

Entropy of a set of examples S with class labels $\{C_1, \dots, C_n\}$:

$$H(S) = - \sum_{i=1}^n p_i \log_2 p_i$$

where p_i is the relative frequency (probability) of class C_i .

- **Information Gain** measures reduction in entropy when a feature is used to split a set into two or more subsets

IG for feature A that splits a set of examples S into $\{S_1, \dots, S_m\}$:

$$IG(S, A) = (\text{original entropy}) - (\text{entropy after split})$$

$$IG(S, A) = H(S) - \sum_{i=1}^m \frac{|S_i|}{|S|} H(S_i)$$

Each subset is weighted in proportion to its size

Tutorial Q1(a)

a) What is the entropy of this data set with respect to the target class label *Result*?

	Name	Hair	Height	Weight	Dublin	Result
1	Sarah	blonde	average	light	no	sunburned
2	Dana	blonde	tall	average	yes	none
3	Alex	brown	short	average	yes	none
4	Annie	blonde	short	average	no	sunburned
5	Emily	red	average	heavy	no	sunburned
6	Pete	brown	tall	heavy	no	none
7	John	brown	average	heavy	no	none
8	Katie	brown	short	light	yes	none

Entropy(Dataset)

$$= -(3/8) * \log_2(3/8) - (5/8) * \log_2(5/8)$$

$$= 0.5306 + 0.4238 = 0.9544$$

$$H(S) = - \sum_{i=1}^n p_i \log_2 p_i$$

Tutorial Q1(b)

b) Construct the decision tree that would be built with Information Gain for this data set. Show your work for selection of the root feature in your tree.

	Name	Hair	Height	Weight	Dublin	Result
1	Sarah	blonde	average	light	no	sunburned
2	Dana	blonde	tall	average	yes	none
3	Alex	brown	short	average	yes	none
4	Annie	blonde	short	average	no	sunburned
5	Emily	red	average	heavy	no	sunburned
6	Pete	brown	tall	heavy	no	none
7	John	brown	average	heavy	no	none
8	Katie	brown	short	light	yes	none

- Steps:**
1. Calculate overall dataset entropy.
 2. Calculate entropy for each feature.
 3. Calculate Information Gain for each feature.

Tutorial Q1(b)

Calculate entropy values for all features:

$$\text{Entropy}(\text{Hair}=\text{blonde}) = \text{Entropy}(1/3, 2/3) = 0.9183$$

$$\text{Entropy}(\text{Hair}=\text{brown}) = \text{Entropy}(0/4, 4/4) = 0$$

$$\text{Entropy}(\text{Hair}=\text{red}) = \text{Entropy}(1/1, 0/1) = 0$$

$$\text{Entropy}(\text{Height}=\text{average}) = \text{Entropy}(1/3, 2/3) = 0.9183$$

$$\text{Entropy}(\text{Height}=\text{tall}) = \text{Entropy}(2/2, 0/2) = 0$$

$$\text{Entropy}(\text{Height}=\text{short}) = \text{Entropy}(1/3, 2/3) = 0.9183$$

$$\text{Entropy}(\text{Weight}=\text{light}) = \text{Entropy}(1/2, 1/2) = 1$$

$$\text{Entropy}(\text{Weight}=\text{average}) = \text{Entropy}(1/3, 2/3) = 0.9183$$

$$\text{Entropy}(\text{Weight}=\text{heavy}) = \text{Entropy}(1/3, 2/3) = 0.9183$$

$$\text{Entropy}(\text{Dublin}=\text{no}) = \text{Entropy}(2/5, 3/5) = 0.9710$$

$$\text{Entropy}(\text{Dublin}=\text{yes}) = \text{Entropy}(3/3, 0/3) = 0$$

Hair	Result
blonde	sunburned
blonde	none
brown	none
blonde	sunburned
red	sunburned
brown	none
brown	none
brown	none

Height	Result
blonde	average
blonde	tall
brown	short
blonde	short
red	average
brown	tall
brown	average
brown	short

Tutorial Q1(b)

Use Information Gain to choose best feature to split for root node. Try each feature in turn...

$$\begin{aligned} \text{IG}(\text{Hair}) &= \text{Entropy}(\text{Dataset}) \\ &\quad - p(\text{Hair}=\text{blonde}) * \text{Entropy}(\text{Hair}=\text{blonde}) \\ &\quad - p(\text{Hair}=\text{brown}) * \text{Entropy}(\text{Hair}=\text{brown}) \\ &\quad - p(\text{Hair}=\text{red}) * \text{Entropy}(\text{Hair}=\text{red}) \\ &= 0.9544 - (3/8)*0.9183 - (4/8)*0 - (1/8)*0 \\ &= 0.610 \end{aligned}$$

$$\text{IG}(\text{Height}) = 0.9544 - (3/8)*0.9183 - (2/8)*0 - (3/8)*0.9183 = 0.2657$$

$$\text{IG}(\text{Weight}) = 0.9544 - (2/8)*1 - (3/8)*0.9183 - (3/8)*0.9183 = 0.0157$$

$$\text{IG}(\text{Dublin}) = 0.9544 - (5/8)*0.9710 - (3/8)*0 = 0.3475$$

$$\begin{aligned} \text{Entropy}(\text{Hair}=\text{blonde}) &= 0.9183 \\ \text{Entropy}(\text{Hair}=\text{brown}) &= 0 \\ \text{Entropy}(\text{Hair}=\text{red}) &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(\text{Height}=\text{average}) &= 0.9183 \\ \text{Entropy}(\text{Height}=\text{tall}) &= 0 \\ \text{Entropy}(\text{Height}=\text{short}) &= 0.9183 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(\text{Weight}=\text{light}) &= 1 \\ \text{Entropy}(\text{Weight}=\text{average}) &= 0.9183 \\ \text{Entropy}(\text{Weight}=\text{heavy}) &= 0.9183 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(\text{Dublin}=\text{no}) &= 0.9710 \\ \text{Entropy}(\text{Dublin}=\text{yes}) &= 0 \end{aligned}$$

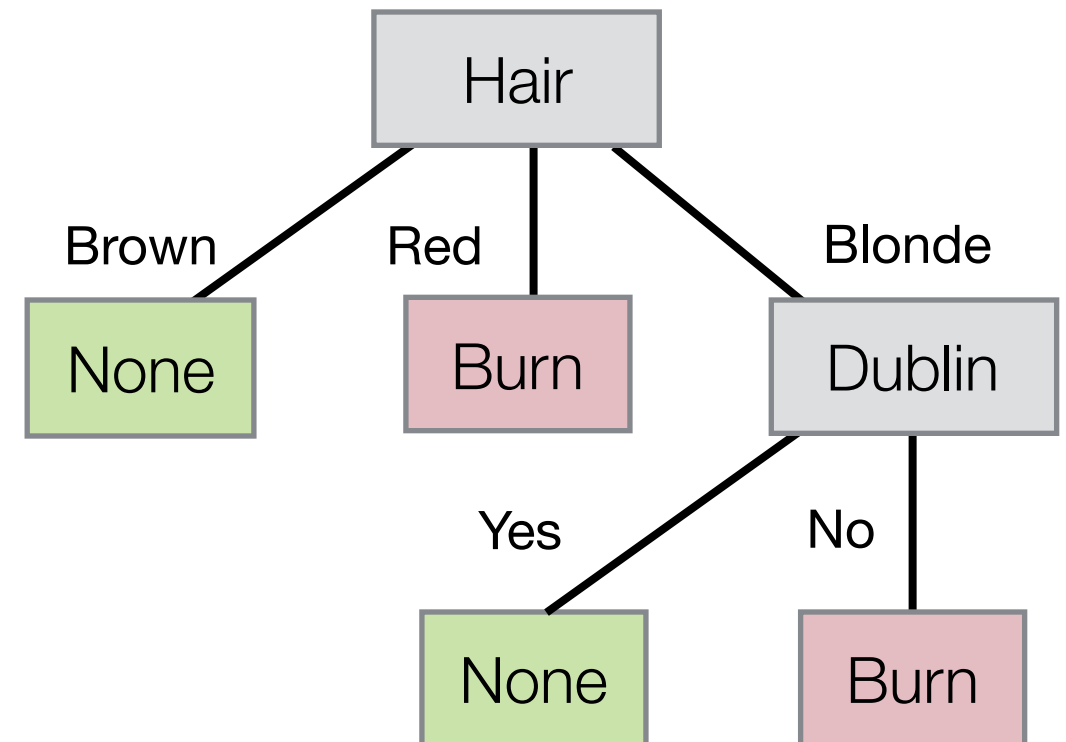
➡ “Hair” will be selected as the feature with the highest IG value.
It perfectly classifies the data for *Hair=brown* & *Hair=red*

Tutorial Q1(b)

- “Hair” selected as the feature with the highest IG value \Rightarrow used to split the root node of the tree.

Child node *Hair=blonde*:

	Name	Hair	Height	Weight	Dublin	Result
1	Sarah	blonde	average	light	no	sunburned
2	Dana	blonde	tall	average	yes	none
4	Annie	blonde	short	average	no	sunburned



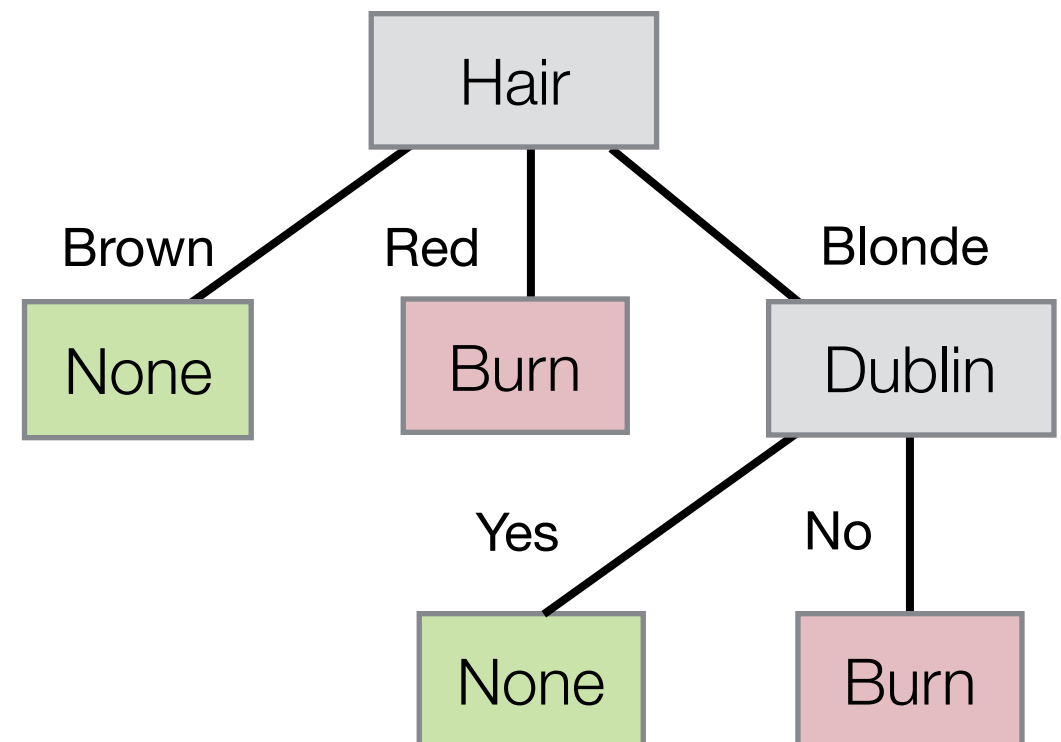
- ➡ The case for *Hair=blonde* contains (2 sunburned, 1 none).
Can split these into pure child nodes using feature “Dublin”.

Tutorial Q1(c)

c) Using your decision tree from (b), how would you classify the following example:

	Hair	Height	Weight	Dublin	Result
X	blonde	average	heavy	no	???

- First, check *Hair=Blonde*
- Next, check *Dublin=No*
- **Output: Sunburned**



Tutorial Q1(d)

- d) Use Naïve Bayes to give the likelihood that the result for the given example is “sunburned”. Then indicate what prediction Naïve Bayes would make. Provide the *probability table* for each of the predicting features.

Probability table for each of the predicting features:

Feature/Value	Class=sunburned	Class=none
Hair=blonde	2/3	1/5
Hair=brown	0/3	4/5
Hair=red	1/3	0/5
Height=average	2/3	1/5
Height=tall	0/3	2/5
Height=short	1/3	2/5
Weight=light	1/3	1/5
Weight=average	1/3	2/5
Weight=heavy	1/3	2/5
Dublin=no	3/3	2/5
Dublin=yes	0/3	3/5
<i>Class Probabilities</i>	<i>3/8</i>	<i>5/8</i>

Tutorial Q1(d)

Use the probability table to calculate the Naïve Bayes scores:

	Hair	Height	Weight	Dublin	Result
X	blonde	average	heavy	no	???

$$v_{NB} = \arg \max_{v_j \in V} P(v_j) \prod_i P(f_i | v_j)$$

Calculate raw probabilities for two classes:

$$P(S) = (2/3) * (2/3) * (1/3) * (3/3) * (3/8)$$

$$P(S) = 0.056$$

$$P(N) = (1/5) * (1/5) * (2/5) * (2/5) * (5/8)$$

$$P(N) = 0.004$$

Normalise probabilities:

$$P(S)' = 0.056 / (0.056 + 0.004) = 0.933$$

$$P(N)' = 0.004 / (0.056 + 0.004) = 0.067$$

Feature/Value	Class=sunburned	Class=none
Hair=blonde	2/3	1/5
Hair=brown	0/3	4/5
Hair=red	1/3	0/5
Height=average	2/3	1/5
Height=tall	0/3	2/5
Height=short	1/3	2/5
Weight=light	1/3	1/5
Weight=average	1/3	2/5
Weight=heavy	1/3	2/5
Dublin=no	3/3	2/5
Dublin=yes	0/3	3/5
<i>Class Probabilities</i>	<i>3/8</i>	<i>5/8</i>

➡ **Output: Sunburned**

Tutorial Q2(a)

a) Which one of the predicting features would be selected by ID3 at the root of a decision tree? Explain your answer.

Example	Credit_History	Debt	Income	Risk
1	bad	low	0to15	high
2	bad	high	15to35	high
3	bad	low	0to15	high
4	unknown	high	15to35	high
5	unknown	high	0to15	high
6	good	high	0to15	high
7	bad	low	over35	moderate
8	unknown	low	15to35	moderate
9	good	high	15to35	moderate
10	unknown	low	over35	low
11	unknown	low	over35	low
12	good	low	over35	low
13	good	high	over35	low
14	good	high	over35	low

Tutorial Q2(a)

a) Which one of the predicting features would be selected by ID3 at the root of a decision tree? Explain your answer.

Calculate entropy for data set:

$$\begin{aligned}\text{Entropy}(\text{Dataset}) &= \\ &= -(6/14) * \log_2(6/14) - (3/14) * \log_2(3/14) \\ &\quad - (5/14) * \log_2(5/14) \\ &= 0.5239 + 0.4762 + 0.5305 \\ &= 1.5306\end{aligned}$$

Calculate entropy values for all features:

$$\text{Entropy}(\text{CH}=\text{bad}) = (1/4) * \log_2(1/4) - (3/4) * \log_2(3/4) = 0.8113$$

$$\text{Entropy}(\text{CH}=\text{unknown}) = -(2/5) * \log_2(2/5) - (1/5) * \log_2(1/5) - (2/5) * \log_2(2/5) = 1.5219$$

$$\text{Entropy}(\text{CH}=\text{good}) = -(1/5) * \log_2(1/5) - (1/5) * \log_2(1/5) - (3/5) * \log_2(3/5) = 1.3710$$

$$\text{Entropy}(\text{Debt}=\text{low}) = -(2/7) * \log_2(2/7) - (2/7) * \log_2(2/7) - (3/7) * \log_2(3/7) = 1.5567$$

$$\text{Entropy}(\text{Debt}=\text{high}) = -(4/7) * \log_2(4/7) - (1/7) * \log_2(1/7) - (2/7) * \log_2(2/7) = 1.3788$$

$$\text{Entropy}(\text{Income}=\text{0to15}) = -(4/4) * \log_2(4/4) = 0$$

$$\text{Entropy}(\text{Income}=\text{15to35}) = -(2/4) * \log_2(2/4) - (2/4) * \log_2(2/4) = 1$$

$$\text{Entropy}(\text{Income}=\text{over35}) = -(1/6) * \log_2(1/6) - (5/6) * \log_2(5/6) = 0.6500$$

	CH	Debt	Income	Risk
1	bad	low	0to15	high
2	bad	high	15to35	high
3	bad	low	0to15	high
4	unknown	high	15to35	high
5	unknown	high	0to15	high
6	good	high	0to15	high
7	bad	low	over35	moderate
8	unknown	low	15to35	moderate
9	good	high	15to35	moderate
10	unknown	low	over35	low
11	unknown	low	over35	low
12	good	low	over35	low
13	good	high	over35	low
14	good	high	over35	low

Tutorial Q2(a)

Use Information Gain to choose best feature to split for root node...

$$\begin{aligned} \text{IG}(\text{CH}) &= \text{Entropy}(\text{Dataset}) \\ &\quad - p(\text{CH}=\text{bad}) * \text{Entropy}(\text{CH}=\text{bad}) \\ &\quad - p(\text{CH}=\text{unknown}) * \text{Entropy}(\text{CH}=\text{unknown}) \\ &\quad - p(\text{CH}=\text{good}) * \text{Entropy}(\text{CH}=\text{good}) \\ &= 1.5306 - (4/14)*0.8113 - (5/14)*1.5219 \\ &\quad - (5/14)*1.3710 \\ &= 0.2656 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(\text{CH}=\text{bad}) &= 0.8113 \\ \text{Entropy}(\text{CH}=\text{unknown}) &= 1.5219 \\ \text{Entropy}(\text{CH}=\text{good}) &= 1.3710 \\ \\ \text{Entropy}(\text{Debt}=\text{low}) &= 1.5567 \\ \text{Entropy}(\text{Debt}=\text{high}) &= 1.3788 \\ \\ \text{Entropy}(\text{Income}=\text{0to15}) &= 0 \\ \text{Entropy}(\text{Income}=\text{15to35}) &= 1 \\ \text{Entropy}(\text{Income}=\text{over35}) &= 0.65 \end{aligned}$$

$$\text{IG}(\text{Debt}) = 1.5306 - (7/14)*1.5567 - (7/14)*1.3788 = 0.0628$$

$$\text{IG}(\text{Income}) = 1.5306 - (4/14)*0 - (4/14)*1 - (6/14)*0.65 = 0.9663$$

➡ “Income” will be selected as the feature to split as it has the highest IG value.

Tutorial Q2(b)

b) What is the main problem with the Information Gain criterion for attribute selection in decision trees?

“Although information gain is usually a good measure for deciding the relevance of an attribute, it is not perfect. **A notable problem occurs when information gain is applied to attributes that can take on a large number of distinct values.** For example, suppose that one is building a decision tree for some data describing the customers of a business. Information gain is often used to decide which of the attributes are the most relevant, so they can be tested near the root of the tree. One of the input attributes might be the customer's credit card number. This attribute has a high mutual information, because it uniquely identifies each customer, but we do not want to include it in the decision tree: deciding how to treat a customer based on their credit card number **is unlikely to generalise to customers we haven't seen before (overfitting).**”

http://en.wikipedia.org/wiki/Information_gain_in_decision_trees

Tutorial Q2(c)

- c) Provide the contingency table of conditional and prior probabilities that would be used by Naïve Bayes to build a classifier for this data set.

Contingency table (probability table) for each of the predicting features:

Risk	high	moderate	low
CH=bad	3/6	1/3	0
CH=unknown	2/6	1/3	2/5
CH=good	1/6	1/3	3/5
Debt=low	2/6	2/3	3/5
Debt=high	4/6	1/3	2/5
Income=0to15	4/6	0	0
Income=15to35	2/6	2/3	0
Income=over35	0	1/3	5/5
<i>Class Probabilities (Priors)</i>	<i>6/14</i>	<i>3/14</i>	<i>5/14</i>

Tutorial Q2(d)

- d) Write down the rule used by Naïve Bayes to classify examples, and apply it to the following example. Which class will be returned by Naïve Bayes for the example?

Example	CH	Debt	Income	Risk
X	bad	low	15to35	???

NB Objective:
$$v_{NB} = \arg \max_{v_j \in V} P(v_j) \prod_i P(f_i | v_j)$$

Calculate raw probabilities for 3 classes, using probability table:

$$P(H) = (3/6) * (2/6) * (2/6) * (6/14) = 0.0238$$

$$P(M) = (1/3) * (2/3) * (2/3) * (3/14) = 0.0317$$

$$P(L) = (0) * (3/5) * (0) * (5/14) = 0$$

Normalise probabilities:

$$P(H)' = 0.0238 / (0.0238 + 0.0317 + 0) = 0.4288$$

$$P(M)' = 0.0317 / (0.0238 + 0.0317 + 0) = 0.5712$$

$$P(L)' = 0$$

➡ **Output:
Moderate**