Decision Tree

$$E(S)=\sum_{i=1}^{c}-p_i\log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)
= Entropy (0.36, 0.64)
= - (0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)
= 0.94

$$E(T,X)=\sum_{c\in X}P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(PlayGolf, Outlook) = **P**(Sunny)***E**(3,2) + **P**(Overcast)***E**(4,0) + **P**(Rainy)***E**(2,3)
= (5/14)*0.971 + (4/14)*0.0 + (5/14)*0.971
= 0.693

Rule-Based Classifiers

Weather data

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Outlook		yes count
	sunny	2/5
	rainy	3/5
temperature		
	hot	0/2
	mild	3/5
	cool	2/3
humidity		
	high	1/5
	normal	4/5
windy		
	TRUE	4/6
	FALSE	1/4

If Humidity = normal and ?
then play = Yes

Outlook		yes count	
	sunny	2/3	
	rainy	2/2	
temperature			
	hot	2/3	
	mild	2/2	
	cool	4/5	
windy			
	FALSE	3/3	Greater Coverage
	TRUE	1/2	

Rule 2:
If Humidity = "normal" and windy = FALSE
then play = Yes

Naïve Bayes Classifier

Likelihood Class Prior Probability

$$P(c\mid x)=\frac{P(x\mid c)P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c\mid \mathbf{X}) = P(x_1\mid c)\times P(x_2\mid c)\times\cdots\times P(x_n\mid c)\times P(c)$$

CASE 1: 2nd child male ?

P(servived=yes | pclass = 2nd , age = child , sex = male)
= Alpha * P(pclass = 2nd | servived = yes) * P(age = child | servived = yes) * P(sex = male | servived = yes) * P(servived = yes)
(H4+1)/(K4+4)*(H10+1)/(I10+2)*(H16+1)/(I16+2)*(K4+1)/(K5+2)
= Alpha * 0.002258411978

P(servived=No | pclass = 2nd , age = child , sex = male)
= Alpha * 0.002473390596

Now,
0.002258*Alpha + 0.009783*Alpha =1
=> Alpha = 211.3359516

Naïve Bayes Text Classifier

$$P(c)=\frac{N_c}{N}$$

words in document	in c
Chinese Beijing Chinese	yes
Chinese Chinese Shanghai	yes
Chinese Macao	yes
Tokyo Japan Chinese	no
Chinese Chinese Chinese Tokyo Japan	?

Bernoulli Model

P(c|d_5) ∝ P(c) · P(Chinese|c) · P(Japan|c) · P(Tokyo|c) ·(1 − P(Beijing|c)) · (1 − P(Shanghai|c)) · (1 − P(Macao|c)) =
3/4 · 4/5 · 1/5 · 1/5 · (1−2/5) · (1−2/5) · (1−2/5) ≈ 0.005
P(-c|d_5) ∝ 1/4 · 2/3 · 2/3 · 2/3 · (1−1/3) · (1−1/3) · (1−1/3) ≈ 0.022
Classifier assigns the test document to c = not-China.

$$P(t\mid c)=\frac{T_{c,t}+1}{\sum_{t\in V}(T_{c,t}+1)}=\frac{T_{c,t}+1}{(\sum_{t\in V}T_{c,t})+|V|}$$

$$P(c\mid d) = \alpha * P(c) * \Pi_{t\in d}\,P(t\mid c)$$

esult: $c_{map}=\mathrm{argmax}_c\ P(c|d)$

Multinomial Model

Priors: P(c) = 3/4 and P(c) = 1/4

Conditional probabilities:
P(Chinese|c) = (5 + 1)/(8 + 6) = 6/14 = 3/7
P(Tokyo|c) = P(Japan|c) = (0 + 1)/(8 + 6) = 1/14

P(Chinese|-c) = (1 + 1)/(3 + 6) = 2/9
P(Tokyo|-c) = P(Japan|-c) = (1 + 1)/(3 + 6) = 2/9

P(c|d_5) ∝ 3/4 · $(3/7)^3$ · 1/14 · 1/14 ≈ 0.0003.
P(-c|d_5) ∝ 1/4 · $(2/9)^3$ · 2/9 · 2/9 ≈ 0.0001.
Thus, the classifier assigns the test document to c = China.

Linear Models

Perceptron

For each misclassified training tuple

$$\mathrm{sign}(\mathbf{w}^T\mathbf{x}^k)\neq y^k$$

$$\mathbf{w}=\mathbf{w}+\eta\cdot y^k\mathbf{x}^k$$

Linear Regression

$$\nabla_E(\mathbf{w})=-\frac{1}{n}\sum_{k=1}^{n}(y^k-\mathbf{w}'\mathbf{x}^k)\mathbf{x}^k$$

```
E = (1/(2*n)) * ( np.sum((y-w@X.T)**2) )
```

$$\mathbf{w}\leftarrow\mathbf{w}-\kappa\nabla_E(\mathbf{w})$$

$$\mathbf{w}\leftarrow\mathbf{w}+\kappa\frac{1}{n}\sum_{k=1}^{n}\left(y^k-\mathbf{w}'\mathbf{x}^k\right)\mathbf{x}^k$$

```
w = w + kappa*( (1/n)*( np.sum( (y-w@X.T).T*X, axis=0, keepdims=True ) ) )
```

Logistic Regression

$$\nabla_E(\mathbf{w})=-\frac{1}{n}\sum_{k=1}^{n}\frac{y^k\mathbf{x}^k}{1+e^{y^k\mathbf{w}^T\mathbf{x}^k}}$$

$$\mathbf{w}\leftarrow\mathbf{w}-\kappa\nabla_E(\mathbf{w})$$