Statistical Models for Visual Detection and Recognition

COMPSCI 527

Today: (discrete probabilities)

• Color features and Matlab
• Joint and conditional probabilities
• Bayes's theorem and the Bayes classifier
% Connected components
cc = bwlabel(trace);
mask = cc==2 | cc==3 | cc==4;
red = img(:, :, 1);
green = img(:, :, 2);
blue = img(:, :, 3);
rgb = [red(mask), ...
green(mask), blue(mask)];

n x 3 array of pixel values
function c = colorToScalar(rgb)
rgb = double(rgb);
denom = sum(rgb, 2);
anz = denom ~= 0;
rgb(anz, :) = rgb(anz, :) ./ (denom(anz) * ones(1, 3));
c = rgb(:, 1) - rgb(:, 2);

• Brightness does not matter
• Yellow $\propto [1 1 0]$
• Orange $\propto [1 0.5 0]$
• Blue does not matter

\[ c = \frac{R - G}{R + G + B} \]
Data feature: $x = \text{bin}(c)$
World state: $w = \{O, L\}$

$$\sum_w \sum_x p(w, x) = 1$$
marginals

\[ p(w) = \sum_x p(w, x) \]

\[ p(x) = \sum_w p(w, x) \]
conditionals

\[ p(x \mid w) = \frac{p(w, x)}{p(w)} \]

\[ p(w \mid x) = \frac{p(w, x)}{p(x)} \]

10 of these \((b = 1, \ldots, 10)\)

2 of these \((f = 1, 2)\)
The Bayes Classifier

- $w = f(x)$: given an image observation $x$, find the world state $w$
- we have $p(w|x)$
- $f(x) = \arg \max_w p(w|x)$

$$p(w \mid x) = \frac{p(w, x)}{p(x)}$$
Classifier with Confidence

- \( f(x) = \arg \max_w p(w|x) \) \hspace{1cm} \text{[Bayes classifier]}
- confidence: some function of \( p(w|x) \):
  maybe \( c(x) = 2 \left[ p(f(x)|x) - 1/2 \right] \) for the binary case
- can say “don’t know” if \( c \) is too small

\[
p(w \mid x) = \frac{p(w, x)}{p(x)}
\]
Noisy Functions

- $f(x)$ is a function that maps each image observation $x$ to a world state $w$
- $p(w|x)$ is a function that maps each image observation $x$ to a distribution over world states $w$
- conditional probabilities are *noisy functions*
oranges?

~oranges == lemons?

not really a binary problem!

how well can we possibly do?
Bayes Error Rate

\[ p(w|x) \]

\[
\begin{array}{cccccccccccc}
0.50 & 1.00 & 0.09 & 0.64 & 0.89 & 0.93 & 0.89 & 0.71 & 0.61 & 0.97 \\
0.50 & 0.00 & 0.91 & 0.36 & 0.11 & 0.07 & 0.11 & 0.29 & 0.39 & 0.03 \\
\end{array}
\]

\[ f(x) = \arg \max_w p(w|x) \]

\[ e = 1 - \sum_x p(f(x), x) = 0.164 \]
Cheating Big Time!

Need: training set ∩ test set = ∅
Discrete Bayes’s Theorem

\[ p(w, x) = p(w \mid x)p(x) = p(x \mid w)p(w) \]

\[ p(x \mid w) = \frac{p(w \mid x)p(x)}{p(w)} \]

\[ p(w \mid x) = \frac{p(x \mid w)p(w)}{p(x)} \]

\[ p(x) = \sum_{w'} p(x \mid w')p(w') \]
Bayes Example

[From Russel and Norvig, Artificial Intelligence, Prentice Hall 1995]

• One in 20 people have a stiff neck
• One in 50,000 people have meningitis
• Half the people with meningitis have a stiff neck
• If you have a stiff neck, should you worry about meningitis?
\[ p(w \mid x) = \frac{p(x \mid w)p(w)}{p(x)} \]
Convenient Notation Abuse

\[ p(w | x) = \frac{p(x | w)p(w)}{p(x)} \]

Four functions, one name!

\[ p_{W|x}(w, x) = \frac{p_{x|w}(x, w)p_W(w)}{p_X(x)} \]

[Note: \( p(a, b | c, d) = p((a, b) | (c, d)) \)]

[book uses Pr instead of \( p \)]