Bayesian Inference for Vision

Oct. 2, 2018 Bill Freeman, CSAIL, MIT
this file: notes/bayesColor2.tex

1 Introduction

In this chapter, we discuss probabilistic models for vision. Let the image or video we observe be $\vec{y}$. Let the underlying explanation, the scene, be called $\vec{x}$. We seek to estimate $\vec{x}$, given $\vec{y}$.

Oftentimes, the underlying scene explanation is not completely determined by the observation. Here are some examples:

- Necker cube
- moon crater illusion
- size ambiguity
- illumination strength ambiguity.

Note that our brains are good at hiding any uncertainty from us. We never see the Necker cube as some linear combination of the two possible explantions; it's always either one interpretation or the other.

Because of the ambiguities of the explanations of the data, we can use a probabilistic framework to describe and solve the problems. In particular, we can describe the task of perception through Bayesian Decision Theory [1].

Let's take one of these ambiguous perception problems and work through it as a running example in this chapter. We can simulate the illumination estimation problem as the following.

- Show distribution of surfaces and illuminant.

To make a demonstration using these stimuli, we need 10 volunteers, 9 to draw reflectance values and one to sample the illumination strength. We do that, and then record the simulated observation brightnesses.

- Show reflectance patch values
- Show illumination values
- Show observed brightnesses.

The brightnesses could be, for example:

```
4, 30, 16, 23, 5, 45, 49, 46, 8
```

2 The computation

Let's start by finding $P(\vec{x}|\vec{y})$. Finding this probability will not be the full story–we'll still need to make a decision–but finding $P(\vec{x}|\vec{y})$ is an important first step.

Typically, we know how an image is formed, and so we know the conditional probability distribution, but in the other direction: the probability of the observations, given the scene, $P(\vec{y}|\vec{x})$.

Why is this forward model written as a probability? Because, in general, there will be noise in the observations. Suppose the computer graphic rendering of our scene is $\vec{f}(\vec{x})$. Then, to derive $P(\vec{y}|\vec{x})$, we can start from

$$\vec{y} = \vec{f}(\vec{x}) + \vec{n} \tag{1}$$

where $\vec{n}$ is an additive noise vector. For the case when that is zero-mean Gaussian noise of variance σ, we have $\vec{n}\tilde{}N(0, \sigma)$. Then we have

$$\begin{aligned} P(\vec{y}|\vec{x}) &= N(\vec{y} - \vec{f}(\vec{x}), \sigma^2) & (2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|\vec{y} - \vec{f}(\vec{x})|^2}{2\sigma^2}) & (3) \end{aligned}$$

How do we go from $P(y|x)$ to $P(x|y)$? (omitting vector symbols for brevity now). We use Bayes rule. Because, by fundamental theorems of probability,

$$\begin{aligned} P(y|x) &= P(x|y)P(y) & (4) \\ &= P(y|x)P(x) & (5) \end{aligned}$$

Thus, we have Bayes Rule:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (6)$$

The first term, $P(y|x)$ is called the *likelihood*. It is the likelihood of the data you obseved, given some underlying scene. $P(x)$ is called the *prior probability*. It is what you believe the world is like before you make your measurement of it. $P(y)$ is a normalizing constant that doesn't depend on the scene and is called the *evidence*.

As a warm-up, let's first consider a very toy estimation problem: given just one observation, y_1, the brightness of one of the nine patches of grey, estimate both the brightness of the underlying patch, x_1, and the illumination strength, x_L.

Since the rendering function is $f(x) = x_1 x_L$, then the likelihood function is

$$P(y_1|x_1, x_L) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|\vec{y_1} - x_1 x_L|^2}{2\sigma^2}) \quad (7)$$

- Show the likelihood ridge.

2.1 Shape of likelihood ridge for illumination estimation problem

3 Ways to break ties

- priors
- choose the loss function appropriately for the task.
- marginalize over nuisance variables

4 Priors

For many vision problems, priors can be very helpful. Here are some situations where the prior saves the day and allows for almost always the correct interpretation.

- Show moon craters
- Show clinton/gore
- show hollow mask illusion

For our problem of estimating the illumination strength and patch brightnesses, our priors are given as follows: For the illumination strength, x_L:

$$P(x_L) = \{ \begin{array}{cc} 0.01 & 1 \leq x_L \leq 100 \\ 0 & \text{otherwise} \end{array} \quad (8)$$

For the reflectance of the ith patch, x_i:

$$P(x_i) = \{ \begin{array}{cc} 0.1 & 0 < x_L \leq 1 \\ 0 & \text{otherwise} \end{array} \quad (9)$$

5 The posterior

Given the priors and the likelihood function, we can write the *posterior* probability. It is named posterior because it is the probability of the x parameters *after* the measurements have been made. Combining Eqs. (8), (9), (7), we have the posterior, $P(x|y)$. If the observation noise for each grey patch is assumed to be independent of that of each other, then their joint probability is the product of their individual probabilities. Thus, the likelihoods for each patch just multiply together to give their joint likelihood, giving:

$$P(x_1 \ldots x_N, x_L|\vec{y}) \quad = \quad \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|y_i - x_i x_L|^2}{2\sigma^2}) P(x_i) P(x_L) \tag{10}$$

6 Loss functions

To make a decision, we need to specify the cost of guessing wrong, $L(x, x')$, where x is what we guessed, and x' is what the scene really was. To make a decision, we can minimize the expected loss, selecting $\hat{x}$:

$$\hat{x} = \text{argmin}_x \int dx' P(x'|y) L(x, x') \tag{11}$$

- MAP: The name of the "Maximum A Posteriori" (MAP) estimate is a description of the decision rule used to find it: you select the scene parameters, x, which yield the maximum of the posterior probability. In the Bayesian decision theory framework, this is equivalent to using a "delta function loss".
- MMSE stands for "Minimum Mean Squared Estimate" and corresponds to a squared error loss function in the Bayesian decision theory framework.
- MLM: This is a non-standard estimator, proposed in [2]. This selects the scene parameters x which maximize the integral of the probability density in some local region.

7 Marginalization over nuisance variables

Intuitively, we want integrate over the likelihood ridge in order to be sensitive to the differing thicknesses of the likelihood ridge.

So let's use the joint posterior over the patch reflectances and the lighting strength to find the marginal probability for just the lighting strength. We have

$$\begin{aligned} P(x_L|\vec{y}) &= \int P(x_1 \ldots x_N, x_L|\vec{y}) dx_1 \ldots dx_N & (12) \\ &= P(x_L) \int \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|y_i - x_1 x_L|^2}{2\sigma^2}) P(x_i) dx_1 \ldots dx_N & (13) \\ & & (14) \end{aligned}$$

The integrals over $dx_{1 \ldots N}$ are all independent of each other, so let's consider the integral over dx_i.

$$\begin{aligned} P(x_L|\vec{y}) &= P(x_L) \int \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{|y_i - x_i x_L|^2}{2\sigma^2}) P(x_i) dx_i & (15) \\ &= P(x_L) \frac{1}{\sqrt{2\pi\sigma^2}} \prod_i \int_0^1 \exp(-\frac{|y_i - x_i x_L|^2}{2\sigma^2}) dx_i & (16) \\ & & (17) \end{aligned}$$

We can assume that the observation noise, σ, is very small compared with the interval $0 < x_i < 1$. Then all that matters is whether the interval $0 < x_i < 1$ contains a feasible value for x_i, given the observation y_i. Ie, is there a feasible value for x_L for which $y_i - x_i x_L = 0$ within the interval $0 < x_i < 1$? If yes, then, for small σ, the integral $\int_0^1$ of the Gaussian will approximately equal the integral $\int_{-\infty}^{\infty}$ of the Gaussian. And if no, the integral will approximately equal zero.

Using $\int_{-\infty}^{\infty} \exp(-a(x+b)^2)dx = \sqrt{\frac{\pi}{a}}$, we have

$$
\begin{aligned}
P(x_L|\vec{y}) &= P(x_L)\frac{1}{\sqrt{2\pi\sigma^2}}\prod_i \left\{ \begin{array}{cc} \frac{1}{x_L} & \begin{array}{c} 0 < x_i < 1 \\ \text{or} \\ 0 < \frac{y_i}{x_L} \leq 1 \\ \text{or} \\ y_i < x_L \text{for all } y_i \end{array} \\ 0 & \text{otherwise} \end{array} \right. & (18) \\
&= \frac{0.01}{\sqrt{2\pi\sigma^2}(x_L)^N} \left\{ \begin{array}{cc} 1 & x_L > \max_i y_i \\ 0 & \text{otherwise} \end{array} \right. & (19)
\end{aligned}
$$

where the last line follows from the previous by considering for which values of x_L will $\frac{y_i}{x_L}$ be ≤ 1.

If the largest observed intensity was 49, as in the examples earlier in this chapter, then the marginal posterior probability for the illuminant, x_L, will be proportional to $\frac{1}{x_L^9}$ over the domain 49 to 100, as shown in the figure:

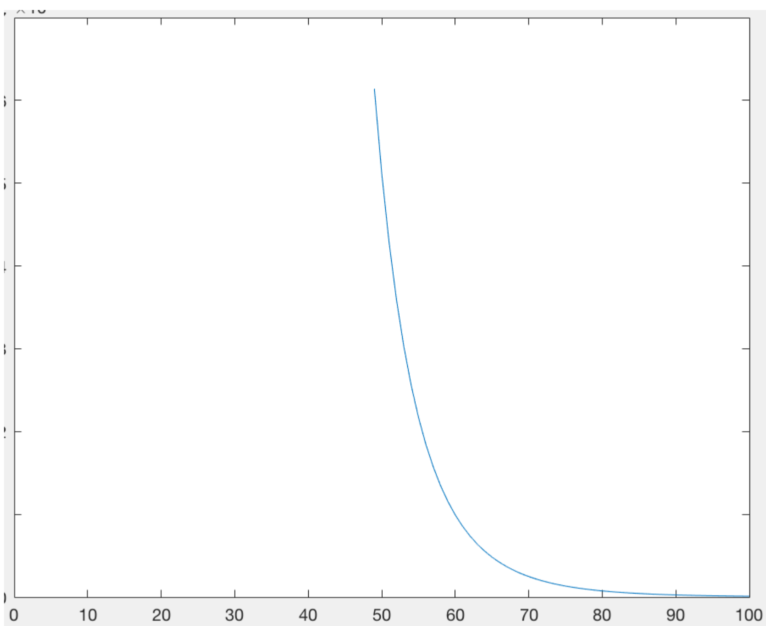

Figure 1: Posterior marginal probability of x_L for the maximum value of y, $y_i = 49$.

References

[1] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.

[2] W. T. Freeman and D. H. Brainard. Bayesian decision theory, the maximum local mass estimate, and color constancy. Technical Report 94-23, Mitsubishi Electric Research Labs., 201 Broadway, Cambridge, MA 02139, 1994.