Lecture 11.

Note Title 4/28/2011

Kalman Filters. and Bayes-Kalman

This is the first of a series of lectures.

Suppose you want to estimate the position x of an object. You have observations $y_1, .. y_m$ drawn (sampled) from a distribution: $P(y|x)$.

How to combine these observations to estimate x?

One strategy: Maximum Likelihood (ML) estimator:

$$\hat{x} = \underset{x}{\mathrm{ARG\,MAX}} \prod_{i=1}^{m} p(y_i|x)$$

If prior knowledge $P(x)$ is known about x, then use Maximum A Posteriori (MAP) estimator:

$$\hat{x} = \underset{x}{\mathrm{ARG\,MAX}}\ P(x|y_1..y_m) = \underset{x}{\mathrm{ARG\,MAX}}\ P(x)\prod_{i=1}^{m} p(y_i|x).$$

Example: If $P(y|x)$ is Gaussian — $P(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}}$.

then ML reduces to $\hat{x} = \frac{1}{m}\sum_{i=1}^{m} y_i$

Note: these estimators are similar to the ones we use in this course. The only difference is that the samples $\{y_i\}$ are generated in a different way — in this example the samples are generated by the observation process (i.e. the world does the sampling for us)

These estimators can be biased, or unbiased. The variance of the estimates will behave like $O(1/m)$ where m is the number of samples.

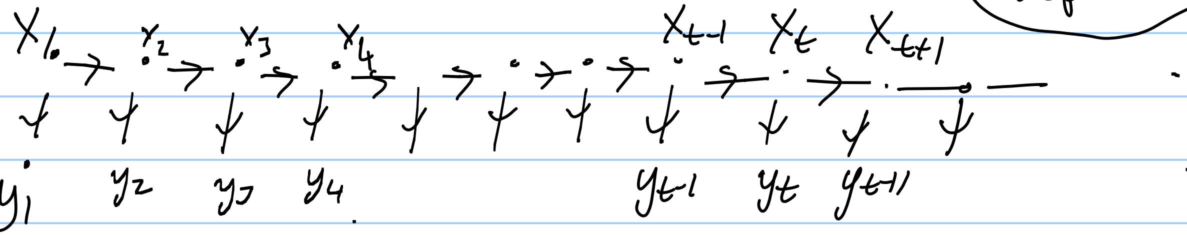
page 2)

But what if the target moves? Suppose X changes over time $X_1, X_2, \ldots X_{t-1}, X_t, X_{t+1}, \ldots$

Note Title 4/23/2006

This requires adding a prior model $P(X_{t+1} | X_t)$ for how the target moves in time – in addition to an observation model $P(y_t | X_t)$ and prior $p(X_1)$

This can be represented graphically: Note: no closed loops.



This gives the Bayes-Kalman model (Kalman is a special case if all distributions are Gaussian).

Let $Y_t = \langle y_t, y_{t-1}, \ldots y_1 \rangle$ be all the obs. up to time t. Then our goal is to estimate $p(X_t | Y_t)$ – the prob of state X_t at time t conditioned on all obs. Y_t up to time t.

We want a recursive algorithm which updates $p(X_t | Y_t)$ to $p(X_{t+1} | Y_{t+1})$ at time $t+1$ with new observation y_{t+1}

$(Y_{t+1} = (y_{t+1}, Y_t))$

Two stages:

(i) prediction. $p(X_{t+1} | Y_t)$

(ii) Correction for new observation $p(X_{t+1} | Y_{t+1})$.

Tracking Airplanes, Spacecraft,

Note: these formulae follow from the graph structure (i.e. the dependencies between variables).

Prediction

(1) $$p(x_{t+1} | Y_t) = \sum_{x_t} p(x_{t+1} | x_t)\, p(x_t | Y_t)$$ (or $\int dx_t$ if x_t is continuous)

Correction.

(2) $$p(x_{t+1} | Y_{t+1}) = \frac{p(y_{t+1} | x_{t+1})\, p(x_{t+1} | Y_t)}{p(y_{t+1} | Y_t)}$$

where $$P(y_{t+1}/Y_t) = \sum_{x_{t+1}} p(y_{t+1} | x_{t+1}) p(x_{t+1} | Y_t)$$ (or $\int dx_{t+1}$) is the normalization term.

Problems: these equations may be very computationally expensive / impractical.

There is a very important special case where these update equations are extremely easy to perform. This is the Kalman Filter. It requires all the distributions to be Gaussian. $p(x_{t+1} | x_t)$, $p(y_t | x_t)$, $p(x_1)$.

We illustrate it in 1-dimension (easy to extend to N-dimension)

$$p(y_t | x_t) = \frac{1}{\sqrt{2\pi}\,\sigma_m} e^{-(x_t - y_t)^2 / 2\sigma_m^2}$$ Gaussian.

$$p(x_{t+1} | x_t) = \frac{1}{\sqrt{2\pi}\,\sigma_p} e^{-(x_{t+1} - x_t - \mu)^2 / 2\sigma_p^2}$$ Gaussian.

$$p(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-(x_1 - \mu_1)^2 / 2\sigma_1^2}$$

<u>Claim:</u> The distribution $P(x_t | Y_t)$ is Gaussian: $p(x_t | Y_t) \sim N(\mu_t, \sigma_t)$. The parameters μ_t, σ_t can be computed recursively by the algebraic update equations below.

<u>First</u>: the prediction $p(x_{t+1} | Y_t)$ is Gaussian.

$$p(x_{t+1} | Y_t) = \int dx_t \, p(x_{t+1} | x_t) \, p(x_t | Y_t)$$

$$= \int_{-\infty}^{\infty} dx_t \frac{1}{\sqrt{2\pi}\,\sigma_p} e^{-(x_{t+1}-x_t-\mu)^2 / 2\sigma_p^2} \frac{1}{\sqrt{2\pi}\,\sigma_t} e^{-\frac{(x_t-\mu_t)^2}{2\sigma_t^2}}$$

$$= \frac{1}{\sqrt{2\pi}\,(\sigma_p^2+\sigma_t^2)^{1/2}} e^{-(x_{t+1}-\mu-\mu_t)^2 / 2(\sigma_p^2+\sigma_t^2)}$$

<u>Second</u>: the new posterior distribution $p(x_{t+1} | Y_{t+1})$ is also Gaussian

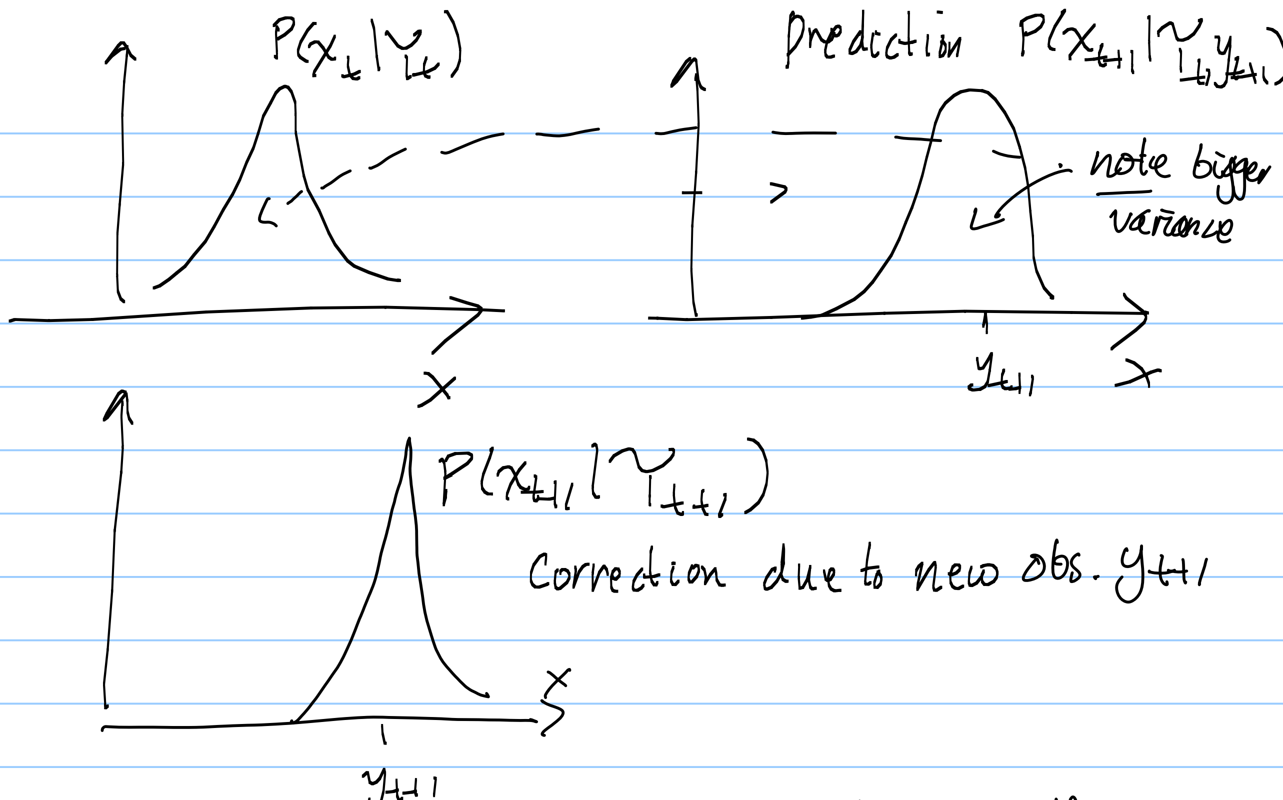
$$p(x_{t+1} | Y_{t+1}) \simeq \frac{p(y_{t+1} | x_{t+1}) \, p(x_{t+1} | Y_t)}{p(Y_t)}$$

$$p(x_{t+1} | Y_{t+1}) = N(\mu_{t+1}, \sigma_{t+1})$$

<u>Kalman's Update Equations:</u>

$$\mu_{t+1} = \underbrace{\mu + \mu_t}_{\text{prediction}} - \frac{(\sigma_t^2+\sigma_p^2)}{\sigma_m^2 + (\sigma_t^2+\sigma_p^2)} \underbrace{\{(\mu+\mu_t) - y_{t+1}\}}_{\text{correction.}}$$

$$\sigma_{t+1}^2 = \frac{\sigma_m^2(\sigma_t^2+\sigma_p^2)}{\sigma_m^2 + (\sigma_t^2+\sigma_p^2)}$$



The Kalman Update equations — Kalman Filter — reduces Bayes-Kalman to algebraic equations to update the means and covariances of the Gaussians. (Note, if we know the means and covariances then we know the distribution).

But Kalman relies on Gaussian distributions and many important real-world applications are non-Gaussian.

Some Examples: (1.) Tracking an object, and another object appears nearby. (next lecture).

(2.) x_t, y_t take finite set of values (e.g. Biology Applications)

Special Cases of Kalman Filter

Case (1) Suppose $\sigma_m = 0$, i.e. the observation are perfect. Then:

(a) $\mu_{t+1} = (\mu + \mu_t) - (\mu + \mu_t) + y_{t+1} = y_{t+1}$

Best estimate is current observation (ignore past obs.).

(b) $\sigma_{t+1} = 0$.

We know position of state with perfect precision.

Case (2) Suppose $\sigma_p = 0$, i.e. we have perfect prediction.

Then:

(a) $$\mu_{t+1} = \frac{\sigma_t^2}{\sigma_m^2 + \sigma_t^2} y_{t+1} + \frac{\sigma_m^2}{\sigma_m^2 + \sigma_t^2}(\mu + \mu_t)$$ weighted average

(b) $$\sigma_{t+1}^2 = \frac{\sigma_m^2 \sigma_t^2}{\sigma_m^2 + \sigma_t^2}$$

If we also have $\mu = 0$ (so X_t is constant)

then $$\mu_{t+1} = \frac{\sigma_t^2}{\sigma_m^2 + \sigma_t^2} y_{t+1} + \frac{\sigma_m^2}{\sigma_m^2 + \sigma_t^2} \mu_t$$

This reduces to the MAP estimate of a static target that we described at the start of the lecture - It gives a recursive, online, method for estimating μ which can be updated as new data arrive. (Here μ corresponds to the estimate of X).