

Basic Introduction to the Kalman Filter

Dan Kersten *Supplement to Introduction to Neural Networks Psy 5038W*

Dec 9, 2009

Please do not cite without permission

Kalman filter: Bayesian background

Assume we are given states y_t that evolve with time, together with a set of measurements x_t of those states. We have a model of how the states evolve given past states, $p(y_{t+1}|y_t, \dots)$, but we can't observe them directly. We can make measurements that are related, but noisy: $p(x_t, \dots|y_t)$. We make two specific assumptions:

Markov assumption—only immediate past matters:

$$p(y_{t+1}|y_0, y_1, \dots, y_t) = p(y_{t+1}|y_t)$$

Measurements are conditionally independent given y_{t+1}

$$p(x_{t+1}, x_t, x_{t-1}, \dots|y_{t+1}) = p(x_{t+1}|y_{t+1})p(x_t|y_{t+1})p(x_{t-1}|y_{t+1})\dots$$

Prediction step:

$$\begin{aligned} p(y_{t+1}|x_0, x_1, \dots, x_t) &= \sum_{y_t} p(y_{t+1}, y_t|x_0, x_1, \dots, x_t) \\ &= \sum_{y_t} p(y_{t+1}|y_t)p(y_t|x_0, x_1, \dots, x_t) \end{aligned}$$

The result, $p(y_{t+1}|x_0, x_1, \dots, x_t)$, is used in the next,

Measurement/Correction step:

where we update the posterior with a new measurement x_{t+1}

$$p(y_{t+1}|x_0, x_1, \dots, x_t, x_{t+1}) = \frac{p(x_{t+1}|y_{t+1})p(y_{t+1}|x_1, x_2, \dots, x_t)}{\sum_{y_{t+1}} p(x_{t+1}|y_{t+1})p(y_{t+1}|x_1, x_2, \dots, x_t)}$$

$p(y_{t+1}|x_0, x_1, \dots, x_t, x_{t+1})$ provides the updated distribution which becomes the “next $p(y_t|x_0, x_1, \dots, x_t)$ ” to use on the right hand side of the prediction step.

Kalman filter: Simple 1D example

We can build our intuitions with a simple concrete 1D example, inspired by Max Welling's tutorial¹. Imagine a ship that at time $t = 0$ starts from harbor at position $y = y_0$. It sets off with constant velocity, gaining distance c at each time step, so we could predict its position into the future as: $y_{t+1} = y_t + c$. But there could be some uncertainty due to various factors, such as buffeting by waves. Thus we can write a generative model for the prior in terms of the true state plus a gaussian noise term:

$$y_{t+1} = y_t + c + w_t$$

Taking the expectation, $E[\cdot]$ and the variance $\text{var}[\cdot]$, of the above, we have:

$$\hat{y}_{t+1} = \hat{y}_t + c$$

$$\sigma_{t+1}^2 = \sigma_t^2 + \sigma_w^2$$

So if there was no uncertainty, $\sigma_w = 0$, then given an initial state $y_{t=0} = y_0$, we could recursively predict any future value of y precisely. However, if $\sigma_w > 0$, then as time goes on, we would become increasingly uncertain of where the ship is. We'd need some measurement, x , such as from a GPS signal, to tell us where the ship might be. But a typical measurement isn't perfectly accurate either, so with another additive gaussian noise assumption, we write:

$$x_t = y_t + v$$

where v represents the noise. Thus we have two pieces of information—first, a mean based on the prior model, and the second on some data. How should we combine them? We've seen before how to integrate information about separate sources—we add them, weighted by their reliabilities. For conditionally independent pieces of information, with gaussian distributions, the rule is:

$$\hat{m} = \frac{r_1}{r_1 + r_2} \hat{m}_1 + \frac{r_2}{r_1 + r_2} \hat{m}_2$$

where

$$r_i = \frac{1}{\sigma_i^2}$$

¹http://www.ics.uci.edu/~welling/classnotes/papers_class/KF.ps.gz

Exercise refresher: To see this, suppose we are trying to estimate the mean of a distribution, call it y , and assume that we have two noisy measurements m_1 and m_2 of y :

$$m_1 = y + v_1$$

$$m_2 = y + v_2$$

Assuming the measurements are independent given y ,

$$p(m_1, m_2|y) = p(m_1|y)p(m_2|y)$$

and gaussian, we have:

$$= \kappa \prod_i e^{-\frac{(m_i - y)^2}{\sigma_i^2}}$$

Taking the derivative of the log shows that the peak occurs when y is:

$$\frac{\mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2} + \frac{\mu_1 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

This suggests that we can combine our two sources of information about the ship as:

$$\hat{y}'_{t+1} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_{t+1}^2} \hat{y}_{t+1} + \frac{\sigma_{t+1}^2}{\sigma_v^2 + \sigma_{t+1}^2} x_{t+1}$$

What is the uncertainty in \hat{y}'_{t+1} ? Recall that variance of a linear sum of random variables u and v is given by, $\text{var}[\alpha u + \beta v] = \alpha^2 \text{var}[u] + \beta^2 \text{var}[v]$. So we have:

$$\sigma_{t+1}'^2 = \frac{\sigma_{t+1}^2 \sigma_v^2}{\sigma_{t+1}^2 + \sigma_v^2}$$

Let's see if things make sense. Suppose we totally trust the measurement x_t . Total trust means $\sigma_v^2 = 0$. Then $\hat{y}'_{t+1} = x_{t+1}$. On the other hand, suppose we have no confidence in the measurement: $\hat{y}'_{t+1} = \hat{y}_{t+1}$. Although it didn't help that time, in general a measurement "never hurts"-uncertainty either stays the same or goes down.

We can put our equations into a standard form by defining the Kalman gain, K as:

$$K_{t+1} = \frac{\sigma_{t+1}^2}{\sigma_{t+1}^2 + \sigma_v^2}$$

and then writing the update equations for the variance and state as:

$$\sigma_{t+1}'^2 = (1 - K_{t+1})\sigma_{t+1}^2$$

$$\hat{y}'_{t+1} = \hat{y}_{t+1} + K_{t+1}(x_{t+1} - \hat{y}_{t+1})$$

Note that complete trust in the data corresponds to $K = 1$, and no trust to $K = 0$.

Kalman filter: General linear case

We now write down, without proof, the corresponding generalization to vector states, y_t , vector measurements, x_t , and multivariate gaussian noise vectors w_t and v_t . The prior, generative model is given by:

$$y_{t+1} = Ay_t + w_t$$

and the measurement vector by:

$$x_t = Hy_t + v_t$$

The dimensionality of the state vector may be greater than the measurement, and thus the matrix H , would be rectangular. The covariance matrix P_t replaces the variance σ_t^2 , *representing the uncertainty in the state estimates*. Q is the covariance matrix for w replacing σ_w^2 , and R is the covariance matrix for v , replacing σ_v^2 .

Initialization: Assume starting values.

$$y_{t=0} = y_0$$

$$P_{t=0} = P_0$$

Model Prediction: Predict next state and covariance at time $t + 1$ using knowledge from the generative model:

$$\hat{y}'_{t+1} = A\hat{y}_t$$

$$P'_{t+1} = AP_tA^T + Q$$

Note that in general, P incorporates the change in uncertainty brought by a measurement at time t . (If we included the control vector u , and we would have: $\hat{y}'_{t+1} = A\hat{y}_t + Bu$.)

Measurement/Correction equations:

$$K_{t+1} = P'_{t+1}H^T(HP'_{t+1}H^T + R)^{-1}$$

$$\hat{y}_{t+1} = \hat{y}'_{t+1} + K_{t+1}(x_{t+1} - H\hat{y}'_{t+1})$$

$$P_{t+1} = (I - K_{t+1}H)P'_{t+1}$$

Now cycle back to the *Prediction* step using these updated state and covariance estimates for the new time t . Note that we've used the "prime symbol" e.g. P' , to mark the variables differently than in the ship illustration—we use the prime to distinguish the state and covariance variables at each time step in the algorithm.

The Mathematica notebook for Lecture 27 implements the 1D ship example, and a 2D tracking example. The 2D example includes a four dimensional state vector, and 2D measurements.