CS572 Lecture Note

Group 10: Ingyo Chung, Ayesha Sahar, and Si-Hwan Heo

November 5, 2018

Korea Advanced Institute of Science and Technology, KAIST

Table of contents

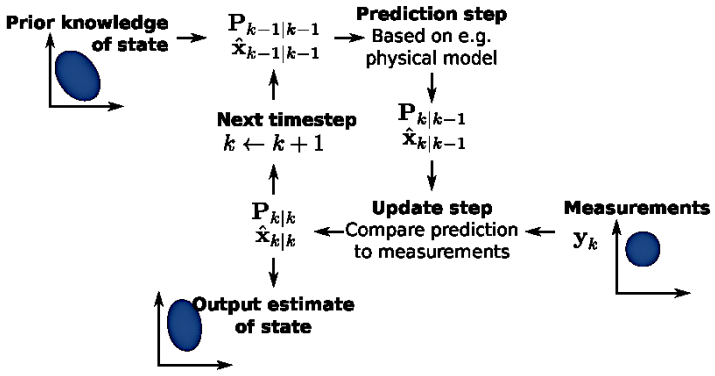
- **Kalman Filter for General System**

An algorithm that uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone [1].

$$x_{k+1} = f(x_k, t) + w_k \quad \approx A_k x_k + w_k$$

$$y_{k+1} = h(x_k, t) + w_k \quad \approx H_k x_k + w_k$$

The below diagram explains the basic steps of Kalman filtering: prediction and update. It also illustrates how the filter keeps track of not only the mean value of the state, but also the estimated variance.



The Kalman filter keeps track of the estimated state of the system and the variance or uncertainty of the estimate. The estimate is updated using a state transition model and measurements. $\hat{x}_{k|k-1}$ denotes the estimate of the system's state at time step k before the k-th measurement y_k has been taken into account; $P_{k|k-1}$ is the corresponding uncertainty.

- **Continuous Kalman Filter**

$$\hat{x} = fx + Gw(t)$$

$$y = Hx + v(t)$$

$$E[w(t)w^T(s)] = Q(\delta(t - S)$$

$$E[v(t)v^T(s)] = R(\delta(t - S)$$

$$E[v(t)w^T(s)] = 0$$

$$E[w] = 0,\ E[v] = 0$$

$$Q_t = Q\Delta t \ \ where \ \ \Delta t \ \ is \ \ the \ \ sampling \ \ interval$$

$$R_t = \frac{R}{\Delta t}$$

$$K_t = P_{t|t-1}H_t^T\left(H_tP_{t|t-1}H_t^T + R_t\right)^{-1}$$

$$= \Delta tP_{t|t-1}H_t^T\left(\Delta tH_tP_{t|t-1}H_t^T + R_t\right)^{-1}$$

$$= \Delta tP_{t|t-1}H_t^TR^{-1} \qquad \Delta t \ll 1$$

$$\cong \Delta tk$$

So

$$P_{t+1|t} = A_tP_tA_t^T + G_tQ_tG_t^T$$

$$= (I + \Delta tF)(I - K_tH_t)P_{t|t-1}(I + \Delta tF)^T + G_tQ_tG_t^T$$

$$O(\Delta t^2) = 0$$

$$P_{t+1|t} = P_{t|t-1} + \Delta tFP_{t|t-1} + \Delta tP_{t|t-1}F^T - \Delta tKH_tP_t + G_t\Delta tGG_t^T$$

$$\Delta t \to 0$$

Riccati equation

$P = FP + PF^T - PH_tR^{-1}HP$ +GQG^T

Thus, in this way we get a Kalman–Bucy filter which is a continuous time version of the Kalman filter.

$$\hat{x} = f\hat{x} + K(y - H\hat{x})$$

where the Kalman gain is given by

$$K = PH^TR^{-1}$$

- **Matrix Fraction Decomposition (A method to solve Riccati equation)**

Consider a Square Matrix M(t)

$$M(t) = A(t)B^{-1}(t)$$

$$\text{B(t)}B^{-1}(t) = I \ -> BB^{-1}$$

The Riccati Differential Equation can be solved by using a technique, called the Matrix Fraction Decomposition.

Consider a square matrix M (t) decomposed into two square matrices A(t) and B(t) ,

$$M(t) = A(t)B^{-1}(t)$$

where B is non-singular and both A and B are differentiable with respect to time t . The above expression is called a fraction decomposition of Matrix M .

Differentiating $B(t)B^{-1}(t) = I$ (identity matrix) with respect to time t,

$$BB^{-1} + BB^{-1} = 0$$

Therefore,

$$\frac{d}{dt}\ B^{-1}(t) = -B^{-1}\frac{d}{dt}\ B(t).B^{-1}$$

Now let us represent the covariance matrix,

$$P\,(t) = A\,(t)\,(B)^{-1}\,(t\,)$$

Hence, it becomes

$$\frac{d}{dt}P(t) = \acute{A}B^{-1} + A\acute{B}^{-1}$$

$$= \acute{A}B^{-1} + AB^{-1}\acute{B}B^{-1}$$

From the Recatti Equation

$$\frac{d}{dt}P(t) = FAB^{-1} + AB^{-1}F^T + AB^{-1}H^TR^{-1}HAB^{-1} + GQG^T$$

Equating the right hand sides of above two equations and post-multiplying B yield.

$$\acute{A} - AB^{-1}\acute{B} = (A + GQG^TB) - AB^{-1}(H^TR^{-1}HA - F^TB)$$

Therefore, if we find A and B that satisfy:

$$\acute{A} = FA + GQG_tB$$
$$\acute{B} = H^TR^{-1}HA - F^TB$$

Then P(t)=A(t) $B^{-1}(t)$ satisfies the Riccati differential equation. Above two equations are Linear differential equations with respect to matrices A and B They can be rearranged as

The Hamiltonian Matrix

$$\frac{d}{dt}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} = \begin{bmatrix} F(t) & G(t)Q(t)G^T(t) \\ H^T(t)R^{-1}(t)H(t) & -F(t) \end{bmatrix}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix}$$

- **Basic of Deep Learning**

Consider a dataset (X, Y). Our goal is to find the weights W such that

y_t=f($w^{-t}X^t$)

Let's define the cost function as

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

Learning as optimization

$$\nabla E(w) = \frac{\partial E}{\partial w1} \dots \dots \dots \dots \frac{\partial E}{\partial wn}$$

This is the gradient descent of steepest ascent.

Gradient Descent Method

This is an optimization technique that is used to decrease any function by repetitively moving towards the descent and the steepest descent is determined by the negative of the gradient.

Learning Rate

The size of the step is called learning rate. Size of the step determines how much points will be covered in each iteration means we can move towards the descent more quickly but in this case, there is a risk of missing the points [2].

Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) is an alternative to standard gradient descent method and it is used for faster convergence [2].

Example # 1

$$OUTPUT = \pm 1$$

$$f(u) = sgn(u)$$

f is significant cycle over examples

$$correct \ \ if \ \ \text{No change}$$

$$if \ \ error \ \ then \ \ \Delta w = yx$$

Perceptron Convergence theorem says that if examples are linearly separable then this is convergence.

Example # 2

$$E(w) =< 1/2 \ \ |w - x|^2 >$$

$$= \frac{1}{m}\sum i \ \ 1/2(w - x_i)^2$$

$$\frac{\partial E}{\partial w} = \frac{1}{m}\sum \quad (\frac{\frac{\partial}{\partial w}}{2}(w - x)^2$$

$$\frac{1}{m}\sum \quad w - x_i) = 0$$

$$w = \frac{1}{m}\sum \quad X_i =< x >$$

$$\Delta w = -\gamma\frac{\gamma\partial}{\partial w}(w - x)^2$$

$$= -\gamma(x - w)$$

$$w(t) = w(t-1) + \frac{1}{t}(x(t) - w(t-1)$$

$$w(1) = 1\big(x(1)\big) = x(1)$$

$$w(2) = w(1) + \frac{1}{2}\big(x(2) - w(1)\big) = \frac{w(1) + x(2)}{2}$$

References

[1]. Wikipedia contributors. (2018, November 1). Kalman filter. In Wikipedia, The Free Encyclopedia. Retrieved 2018, from https://en.wikipedia.org/w/index.php?title=Kalman_filter&oldid=866849029

[2]. Wikipedia contributors. (2018, November 9). Gradient descent. In Wikipedia, The Free Encyclopedia. Retrieved , from https://en.wikipedia.org/w/index.php?title=Gradient_descent&oldid=868096062