IEOR 165 – Lecture 14
Holt-Winters Algorithm

1 Affine Discrete Time Model

An affine discrete time model can be written as

$$x_{t+1} = Ax_t + Bu_t + k,$$

where $x_t \in \mathbb{R}^n$ is a vector of states, $u_t \in \mathbb{R}^q$ is a vector of inputs, and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times q}$, $k \in \mathbb{R}^n$ are matrices/vectors of suitably-defined size. Such models can be used for forecasting. For instance, the state x_t might represent the demand of a basket of goods at time t, and u_t might be economic indicators or other determinants of demand for each item in the basket of goods.

The benefit of these models is that identifying the models and then using them for forecasting is a relatively simple process, consisting of the following steps:

1. Historical data (u_t, x_t) for $t = -n, -n+1, \ldots, 0$ is used to estimate the A, B, k of the model.
2. Different scenarios for the future inputs u_t are generated.
3. The forecast is generated by recursively computing

$$\begin{aligned} x_1 &= Ax_0 + Bu_0 + k \\ x_2 &= Ax_1 + Bu_1 + k \\ &\vdots \end{aligned}$$

   and so on.

In some cases, we may be able to use domain knowledge to exactly specify some of the future inputs.

1.1 Parameter Identification

Identifying the A, B, k is relatively simple because we can pose the problem as a linear regression. In particular, if we let A_j, B_j, k_j denote the j-th row of A, B, k respectively, then we can estimate the model parameters by solving the following least squares problems

$$\begin{bmatrix} \hat{A}_j \\ \hat{B}_j \\ \hat{k}_j \end{bmatrix} = \arg\min \left\| Y^j - X \begin{bmatrix} A_j' \\ B_j' \\ k_j' \end{bmatrix} \right\|_2^2,$$

where we define

$$Y^j = \begin{bmatrix} x^j_{-n+1} \\ x^j_{-n+2} \\ \vdots \\ x^j_0 \end{bmatrix} \qquad X = \begin{bmatrix} x'_{-n} & u'_{-n} & 1 \\ x'_{-n+1} & u'_{-n+1} & 1 \\ \vdots & & \\ x'_{-1} & u'_{-1} & 1 \end{bmatrix}.$$

1.2 Example: Weight Modeling

Suppose we would like to forecast the weight of an individual. To help construct a temporal model, we have asked the individual to weigh their body weight every morning. Suppose the measured weight data (in units of pounds) is

$$\{w_0, w_1, \ldots\} = \{138, 137.9, 137.3, 137.5, 137.1, 137.0\}$$

Q: Identify the parameters of an affine discrete time model given by

$$w_{n+1} = \alpha \cdot w_n + k.$$

Here, the parameter k encompasses the impact of the basal metabolic rate (BMR), average daily caloric intake each data, and average physical activity in a day.

A: For our predictors, we use the data:

$$x_i = \{138, 137.9, 137.3, 137.5, 137.1\}$$

For our response, we use the data

$$y_i = \{137.9, 137.3, 137.5, 137.1, 137.0\}$$

To solve the linear regression problem, we compute

$$\begin{aligned} \overline{x} &= 137.56 \\ \overline{y} &= 137.36 \\ \overline{xy} &= 18895.31 \\ \overline{x^2} &= 18922.87. \end{aligned}$$

Thus, our estimates are

$$\begin{aligned} \hat{\alpha} &= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{18895.31 - 137.56 \cdot 137.36}{18922.87 - (137.56)^2} = 0.59 \\ \hat{k} &= \overline{y} - \hat{\alpha} \cdot \overline{x} = 137.36 - 0.59 \cdot 135.56 = 57.4. \end{aligned}$$

Our final estimated model is

$$x_{n+1} = 0.59 \cdot x_n + 57.4.$$

2 Additive Holt-Winters Algorithm

Suppose our state is $x_t \in \mathbb{R}$, and that we have observed $x_1, \ldots, x_n$ as data. If the future state follows a constant plus trend model, then the prediction h time steps into the future is given by

$$\hat{x}_{t+h} = \hat{a}_t + \hat{b}_t \cdot h,$$

where $\hat{a}_t$ represents the constant term and $\hat{b}_t$ represents the trend. For instance, a prediction one time step into the future is given by

$$\hat{x}_{t+1} = \hat{a}_t + \hat{b}_t.$$

The next question is how to compute these $\hat{a}_t, \hat{b}_t$ values.

Now suppose that the model describing the $\hat{a}_t, \hat{b}_t$ for $t = 2, \ldots, (n-1)$ is given by

$$\begin{aligned}
\hat{a}_{t+1} &= \alpha x_t + (1-\alpha)(\hat{a}_t + \hat{b}_t) \\
\hat{b}_{t+1} &= \beta(\hat{a}_{t+1} - \hat{a}_t) + (1-\beta)\hat{b}_t,
\end{aligned}$$

with initial conditions

$$\begin{aligned}
\hat{a}_2 &= x_2 \\
\hat{b}_2 &= x_2 - x_1.
\end{aligned}$$

Note that these x_t are measured data, and so we can recursively compute $\hat{a}_3$, then $\hat{b}_3$, then $\hat{a}_4$, and so on. The parameters α, β can be chosen to minimize $\sum_{t=3}^{n}(x_t - \hat{x}_{t|t-1})^2$, where $\hat{x}_{t|t-1}$ denotes the prediction at time t using data until time $t-1$ (i.e., $\hat{x}_{t|t-1} = \hat{a}_{t-1} + \hat{b}_{t-1}$).

2.1 Seasonal Algorithm

Suppose the state also displays seasonality with period d. Then if the future state follows a constant plus trend plus seasonality model, then the prediction h time steps into the future is given by

$$\hat{x}_{t+h} = \hat{a}_t + \hat{b}_t \cdot h + \hat{c}_{t+h},$$

where $\hat{a}_t$ represents the constant term, $\hat{b}_t$ represents the trend, and $\hat{c}_t$ represents the seasonal component. Because $\hat{c}_t$ is periodic with period d, we have that

$$\hat{c}_{t+h} = \hat{c}_{(t+h) \text{ mod } d}.$$

To compute these values, we use the recursive equations

$$\begin{aligned}
\hat{a}_{t+1} &= \alpha(x_{t+1} - \hat{c}_{t+1-d}) + (1-\alpha)(\hat{a}_t + \hat{b}_t) \\
\hat{b}_{t+1} &= \beta(\hat{a}_{t+1} - \hat{a}_t) + (1-\beta)\hat{b}_t \\
\hat{c}_{t+1} &= \gamma(x_{t+1}) - \hat{a}_{t+1}) + (1-\gamma)\hat{c}_{t+1-d}.
\end{aligned}$$

for $t = (d+2), \dots, n$. We use the initial conditions

$$\hat{a}_{d+1} = x_{d+1}$$
$$\hat{b}_{d+1} = (x_{d+1} - x_1)/d$$
$$\hat{c}_j = x_j - (x_1 + \hat{b}_{d+1}(j-1)), \text{for } j = 1, \dots, (d+1).$$

2.2 Kalman Filtering

These approaches are special cases of a more general approach known as Kalman Filtering using the internal model principal. One of the benefits of this more general approach is that we can easily incorporate multiple seasonality components occuring at different rates (i.e., monthly, weekly, daily, etc.). This approach is covered in other courses at Berkeley.

3 More Information and References

The material in Section 2 follows that of the course textbook "Introduction to Time Series and Forecasting" by Brockwell and Davis.