

1 모델링이란?

- 모델링이란

$$y_i = f(x_i) + \epsilon_i$$

의 꼴에서 f 의 모양을 결정하는 과정을 의미한다.

- f 의 모양을 결정할때 데이터에 대한 확실한 사전정보가 있는 경우가 있다. 예를들어 " $f(x)$ 는 x 에 선형변환으로 만들어질 수 있다. (즉 $f(x) = \beta_0 + \beta_1 x$)"라는 사실을 알고 있는 경우이다. 이는

$$y_i = f(x_i) + \epsilon_i,$$

와 같은 모델에서 f 가 어떠한 형태를 가질것인지 미리 알고 있다고 생각한다는 말과 같다. 이처럼 f 가 어떤 모양인지 미리 알고 접근하는 방법을 파라메트릭 모델링 이라고 한다.

- 사전정보가 없어서 f 를 어떻게 모델링할지 감이 안 올 수도 있다. 즉 자료를 봤는데 선형의 모양을 가지는지 어떤지 감을 못잡겠는 경우이다. 이것을 바꾸어 말하면 $\{y_i\}$ 가 $\{x_i\}$ 의 어떤 space에 있는지 감을 못 잡겠다는 뜻이다. 혹은 모델링이 귀찮을 수도 있다. 이럴 경우 $f(x)$ 가 x 의 어떤 특정스페이스 \mathcal{A} 의 부분공간에 존재한다고 가정하고 그 특정스페이스 \mathcal{A} 를 할 수 있는 베이스를 선택하여 문제를 풀 수 있다. 가령 예를들면 " $f(x)$ 가 어떤 공간에 있는지 모르겠는데 최소한 비숍스페이스의 부분공간에 있는것 같아" 라고 생각한다면 웨이블릿 베이스를 선택하여 모델링 하는 것이다. 보통 위와 같은 접근법은 무한대의 basis를 활용한다. 많은 수학자들이 "이런식으로 무한개의 basis를 활용하면 특정공간에 있는 어떠한 함수도 표현할 수 있어요" 라는 식의 증명을 많이 해놓았는데 이러한 증명결과들을 적극적으로 활용하는 셈이다. 요렇게 f 를 표현하는데 무한개의 basis를 활용하는 모델링을 semi-parametric modeling 이라고 한다.

- 웨이블릿과 푸리에변환등으로 $f(x)$ 를 추론하는 것이 대표적인 세미파라메트릭 모델링이다.

- 파라메트릭 모델링도 못하겠고 세미파라메트릭 모델링도 못하겠다면 년파라메트릭 모델링을 할 수 있다. 년파라메트릭 모델링은 $f(x)$ 에 대한 어떠한 가정도 필요하지 않다. 예를들면 "어떠한 식으로 표현가능하다" 라든가 (요건 파라메트릭 스타일) 혹은 "최소한 어떠한 공간안에 있는것 같다" 라든가 (요건 세미파라메트릭 스타일) 하는 식의 가정이 필요하지 않다.

- 년파라메트릭(non-parametric) 은 통계학에서 대충 2개의 의미로 쓰인다. 보통 (1) 자료가 특정한 분포에서 나왔다는 가정이 필요없는 경우 (2) 자료가 특정한 모델 혹은 스트럭처에서 생성된다는 가정이 없는 경우를 의미한다(위키피디아 참고). (1) 과 관련된 용어로는 non-parametric statistics, distribution-free, rank, order-statistics 등이 있다. (2) 와 관련된 용어는 non-parametric regression, non-parametric hierarchical Bayesian models, kernel 등이 있다. 여기에서는 모델링과 관련된 내용을 다루므로 (2)의 경우로 한정하자.

- 년파라메트릭은 베이스스를 설계할때 입력표본 x_i 을 사용한다. 따라서 given data 에 따라 basis 가 달라진다. 이러한 특징을 data adaptive 하다 라고 표현한다. 아래는 년파라메트릭 모델링의 예이다.

$$f(x) = e^{-\frac{\|x-x_1\|}{2h^2}} \beta_1 + e^{-\frac{\|x-x_2\|}{2h^2}} \beta_2 + \cdots + e^{-\frac{\|x-x_n\|}{2h^2}} \beta_n$$

위의 식을 보면 입력데이터를 그대로 쓰지 않고 커널(kernel) 을 한번 씩웠는데 이렇게 입력표본에 커널을 씌우는 방식을 커널모델이라고 한다.

- 년파라메트릭 방법중에 커널을 쓰는 방법이 매우 많다. 하지만 모든 년파라메트릭 방법이 반드시 커널을 써야하는 것은 아니다. 예를 들어 simple kriging 같은 경우는 커널을 쓰지 않는 년파라메트릭 방법으로 볼 수 있다.

- 참고로 년파라메트릭 방법에도 모수가 있다. 구체적으로 위의식에서 β_j 가 모수에 해당한다. 년파라메트릭 방법은 모수가 있냐 없냐가 중요한게 아니라 basis 가 given data 의 함수이냐 아니냐가 중요하다. 따라서 내생각에는 년파라메트릭 모델 보다 data-adaptive model 이 더 좋은 표현 인것 같다. 즉 다시말하면 파라메트릭 모델 혹은 세미파라메트릭 모델은 f 의 스트럭처(structure) 가 모수 θ 만의 함수이다. 따라서 $y_i = f(x_i) + \epsilon_i$ 의 꼴에서 f 를

$$f_{\theta}$$

와 같은 형태로 표현가능하다. 하지만 년파라메트릭 모델은 f 의 스트럭처가 모수와 x_i 에 모두 의존한다.

- 실제로 아까 살펴본 커널모델을 다시 살펴보자. 커널을 결정하는 것은 사실 (1) 밴드윗 (2) 중심점 (3) 높이 인데 여기에서

$$f(x) = e^{-\frac{\|x-x_1\|}{2h^2}} \beta_1 + e^{-\frac{\|x-x_2\|}{2h^2}} \beta_2 + \cdots + e^{-\frac{\|x-x_n\|}{2h^2}} \beta_n$$

아까의 모델이 년파라메트릭인 이유는 (2) 중심점을 표현하는 파라미터가 데이터 자체이기 때문이다.

2 Linear model, Nonlinear model

- 모델링이란

$$y_i = f(x_i) + \epsilon_i$$

의 꼴에서 f 의 모양을 결정하는 과정을 의미한다.

- 이때 f 는 베이스와 파라미터로 구성된다. 그런데 이때 f 가 아래와 같이

$$f = \sum basis \times coef$$

꼴의 형태로 표현할 수 있다면 이 모델을 선형모델이라고 하고 그렇지 않으면 비선형모델이라고 한다.

- 일반적으로 커널은 bandwidth 와 height 에 따라서 결정된다. 예를들어

$$f(x) = e^{-\frac{\|x-x_1\|}{2h^2}}\beta_1 + e^{-\frac{\|x-x_2\|}{2h^2}}\beta_2 + \dots + e^{-\frac{\|x-x_n\|}{2h^2}}\beta_n$$

와 같은 모델에서는 밴드윗이 h 이고 높이가 β_i 라고 이해할 수 있다. 만약에 h 가 known 이라 β_i 만을 추정해도 된다고 하자. 이 경우는 베이스와 코이피션트가 선형으로 결합된 형태가 되므로 이 경우에는 커널모델을 선형모델로 해석할 수 있다. 하지만 h 역시 우리가 추정해야 한다면 이 모델은 비선형모델이 된다.

- 선형모델이 아니지만 적당한 변환을 통하여 선형모델로 바꿀수 있는 모델 역시 넓은범위에서는 선형모델로 친다. 이런 모델을 generalized linear model 이라고 한다. 로지스틱이나 로그선형모델등이 이러한 범주에 속한다.

- 많은 사람들이 $f(x)$ 가 x 에 대하여 선형이면 선형모델이라고 생각하는데 이건 사실이 아니다. 이것은 f 가 x 에 대하여 비선형이라도 파라미터 β 에 대하여 선형이라면 선형모델이라고 한다. 중학교 1학년때 배우는 방정식을 보면 기준을 x 로 보느냐 y 로 보느냐에 따라 1차식이 되기도 하고 2차식이 되기도 하는 수식을 보았을 것이다. 예를 들어보자.

$$x^2y = 0$$

은 이 1차식인가 2차식인가? 이것은 x 에 대한 2차식이지만 y 에 대한 1차식이다. 이와 유사하게

$$f(x) = \beta_0 + \beta_1x + \beta_2x^2$$

은 x 에 대하여는 비선형이지만 $\beta = (\beta_0, \beta_1, \beta_2)'$ 에 대해서는 선형이다.

3 Additive model, Multiplicative model, Hierarchical model

- 모델링이란

$$y_i = f(x_i) + \epsilon_i$$

의 꼴에서 f 의 모양을 결정하는 과정을 의미한다.

- 이때 f 는 컴포넌트(component)들의 조합으로 구성된다고 볼 수 있다. 여기에서 컴포넌트란 베이스와 코이피션트가 결합된 형태로 이해하면 된다. 예를들어

$$y_i = f(x_i) + \epsilon_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

는 β_0 와 $\beta_1 x_i$ 2개의 컴퍼넌트가 있다고 볼 수 있다.

- 각 컴퍼넌트들이 더하기로 연결되어 있으면 additive model 이라고 한다. 각 컴퍼넌트들이 곱의 형태로 표현되어 있으면 multiplicative model 이라고 한다. 각 컴퍼넌트간에 하이라키가 있다면 Hierarchical model 이라고 한다. 3개의 분류는 서로 배타적이지는 않다. 예를들면 아래와 같은 모델을 살펴보자.

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_1X_2\gamma + \epsilon$$

이 모델은 어디티브 모델이지만 (각 컴퍼넌트들이 최종적으로는 +로 연결됨) 살짝 멀티플리케이트브 모델 느낌도 있고 (γ 에 대응하는 베이스는 곱의 형태로 되어있음) 하이라키 모델 (X_1X_2 는 2차식이지만 X_1, X_2 는 1차식이고 1차식을 다 알면 2차식을 construct 할 수 있는 구조임) 이다. 이 모델은 딱 잘라서 무슨 모델이라고 말하기 애매하다. 또한 시계열모형중 하나인 garch 와 같은 경우는 어디티브 모델과 멀티플리케이트브 모델 어느쪽으로도 해석하기 애매하다.

4 결론

- 모델들에 대한 용어들은 서로간의 관계가 애매해서 표나 벤다이어그램으로 이해할 생각은 안하는게 좋다.

- 가령 예를들면 푸리에변환, 웨이블릿 변환은 세미파라메트릭 모델이면서 선형모델이다. 그리고 어디티브 모델이다. 또한 scale 에 따라서 계층구조가 존재하므로 하이라키 모델이다. 그리고 LDA(Latent Dirichlet allocation) 와 같은 모델은 하이라키 모델이며 년파라메트릭 모델이다. 로지스틱은 비선형 모델이지만

generalized linear model 이다. 신경망은 파라메트릭 모델이면서 비선형 모델이고 계층모델이다.