3.
Linear Models for Regression

조윤정

Regression의 목표

- 회귀 모델의 목표 : 입력 값을 통해 예측하는 것
- 선형 회귀 모델의 목표 : 선형 함수 이용

input

D차원의
벡터 입력

output

Target 변수
예측!!

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \boxed{\phi_j(\mathbf{x})} \xrightarrow{\phi_0(\mathbf{x}) = 1} y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

input	$\boxed{y(x)}$	output
D차원의 벡터 입력	$\longrightarrow$	Target 변수
관측값		관측값에 따른 Target

Polynomial basis function(다항 기저 함수)

$$\phi_j(x) = x^j$$

Spline function

Y=x^9

↓

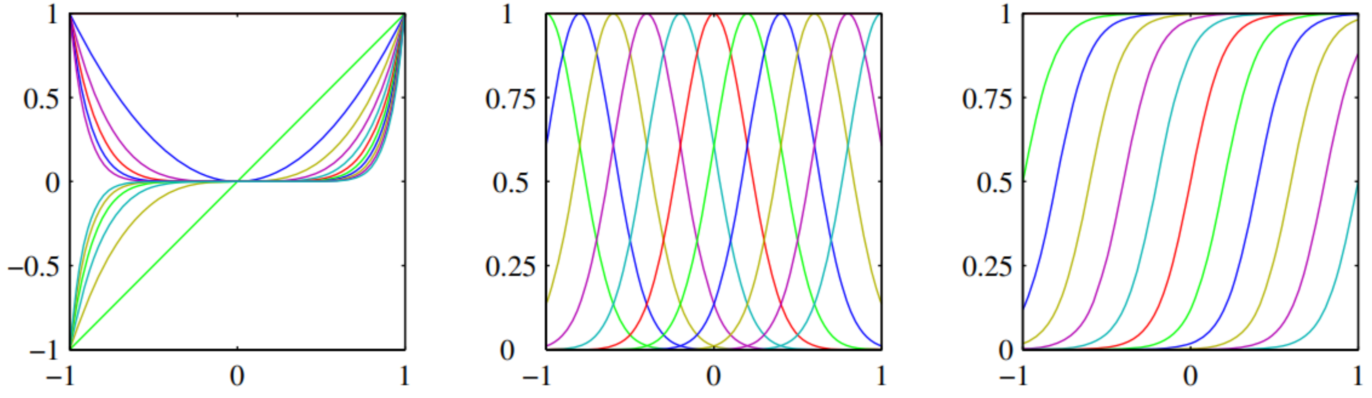
Y=x^2	Y=x^3	Y=x^4	Y=x^5	Y=x^6

Basis function

$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

$$\phi_j(x) = \sigma\left(\frac{x-\mu_j}{s}\right)$$

Simoidal basis function



Polynomial Gaussian simoidal

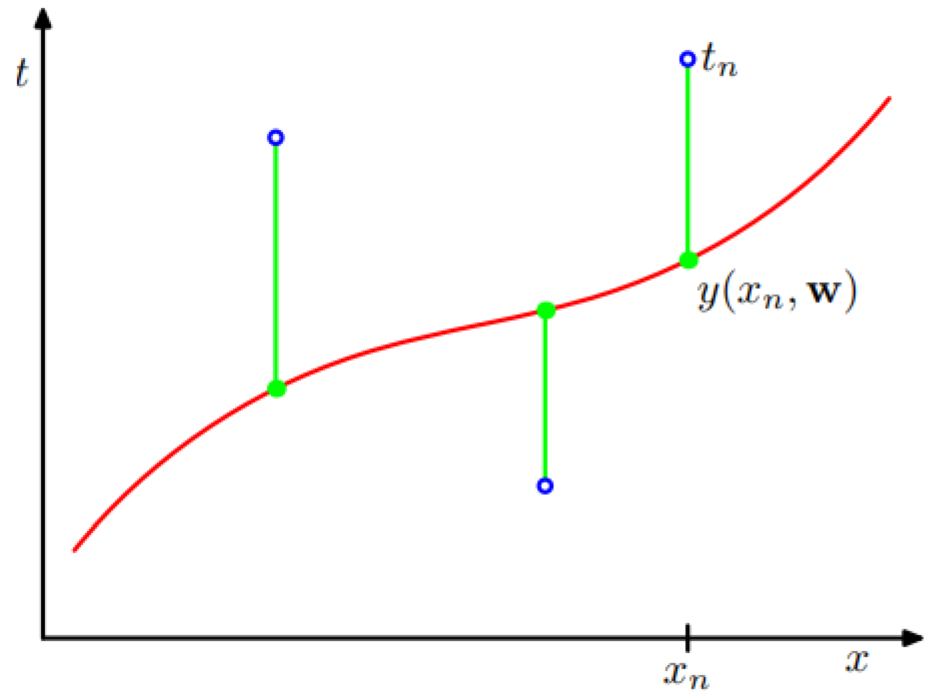
$$\sigma(a) = \frac{1}{1+\exp(-a)}.$$

Logistic sigmoid function

tanh a = 2σ(a) - 1

3.1.1 Maximum likelihood and least squares

“가정”



$$t = \underbrace{y(\mathbf{x}, \mathbf{w})}_{\text{Deterministic function}} + \underbrace{\epsilon}_{\text{Gause noise}}$$

$= 0$

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1}).$$

3.1.1 Maximum likelihood and least squares

$$\mathbb{E}[t|\mathbf{x}] = \int tp(t|\mathbf{x})\,\mathrm{d}t = y(\mathbf{x}, \mathbf{w}).$$

t의 조건부 평균

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1})$$

T가 분포에 자체에 영향 받지 않고 독립적으로 추출되었다면…

$$\begin{aligned}\ln p(\mathbf{t}|\mathbf{w}, \beta) &= \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1}) \\ &= \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \beta E_D(\mathbf{w})\end{aligned}$$

로그를 취해줌

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2.$$

Error function 정리

3.1.1 Maximum likelihood and least squares

$$\nabla \ln p(\mathbf{t}|\mathbf{w}, \beta) = \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}.$$

$$0 = \sum_{n=1}^{N} t_n \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} - \mathbf{w}^{\mathrm{T}} \left(\sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right).$$

$$\mathbf{w}_{\mathrm{ML}} = \left(\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t}$$

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{ t_n - w_0 - \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}_n) \}^2.$$

정리한 결과

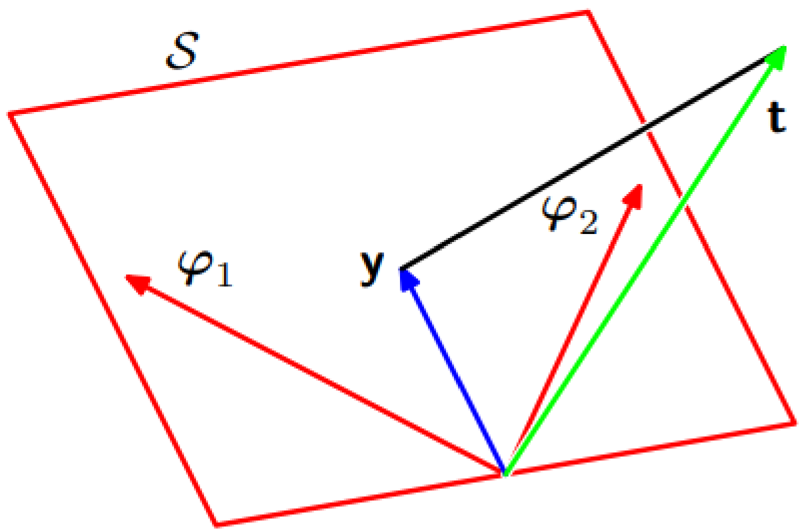
$$w_0 = \bar{t} - \sum_{j=1}^{M-1} w_j \overline{\phi_j}$$

$$\bar{t} = \frac{1}{N} \sum_{n=1}^{N} t_n,$$

$$\overline{\phi_j} = \frac{1}{N} \sum_{n=1}^{N} \phi_j(\mathbf{x}_n).$$

$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N} \sum_{n=1}^{N} \{ t_n - \mathbf{w}_{\mathrm{ML}}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \}^2$$

3.1.2 Geometry of least squares



3.1.3 Sequential learning

$$y(\boxed{\mathbf{x},}\,\mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

같음

$$E = \sum_n E_n \qquad \mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E_n$$

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \eta(t_n - \mathbf{w}^{(\tau)\mathrm{T}}\phi_n)\boxed{\phi_n}$$

Least mean square(LMS)

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boxed{\phi(\mathbf{x}_n)}\}^2.$$

Stochastic gradient descent : 확률적 경사 하강법
Sequential gradient descent : 순차적 경사 하강법

3.1.4 Regularized least squares

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

$$E_W(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}.$$

$$E_D(\mathbf{w}) \cancel{E(\mathbf{w})} = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2$$

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \boxed{\mathbf{w}^{\mathrm{T}}}\phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}.$$

$$\mathbf{w} = \left(\lambda\mathbf{I} + \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}.$$

3.1.4 Regularized least squares

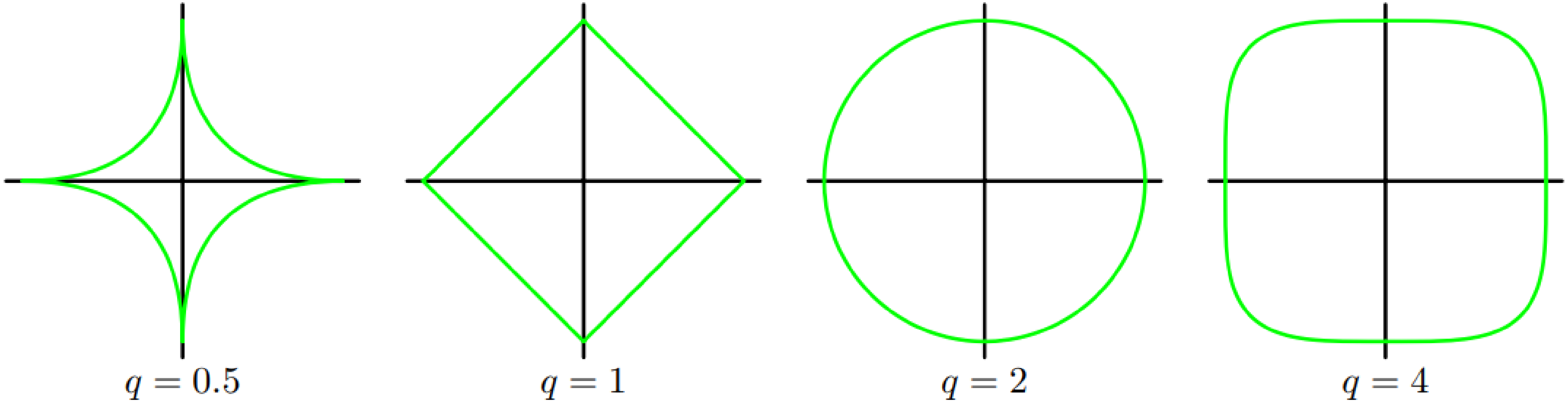


Figure 3.3 Contours of the regularization term in (3.29) for various values of the parameter q.

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$$

일반적인 형태의 정규화 항

3.1.5 Multiple outputs

3.2 The Bias-Variance Decomposition

1장

복잡한 모델 → 최소 제곱법 이용 시 심각한 overfitting 문제

Basis function을 제한한다면? → 모델 유연성에 제약

정규화항 이용 → parameter가 많아도 과적합 조절 가능, 하지만 정규화 계수의 값을 적절히 조절할 수 있어야 함.

베이지론 방법론을 바탕으로 각각의 매개변수를 주변화하면 걱정 X

그 전에 빈도주의 관점의 bias variance trade off 살피기

3.2 The Bias-Variance Decomposition

$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - h(\mathbf{x})\}^2 \, p(\mathbf{x}) \, \mathrm{d}\mathbf{x} + \int \{h(\mathbf{x}) - t\}^2 p(\mathbf{x}, t) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t.$$

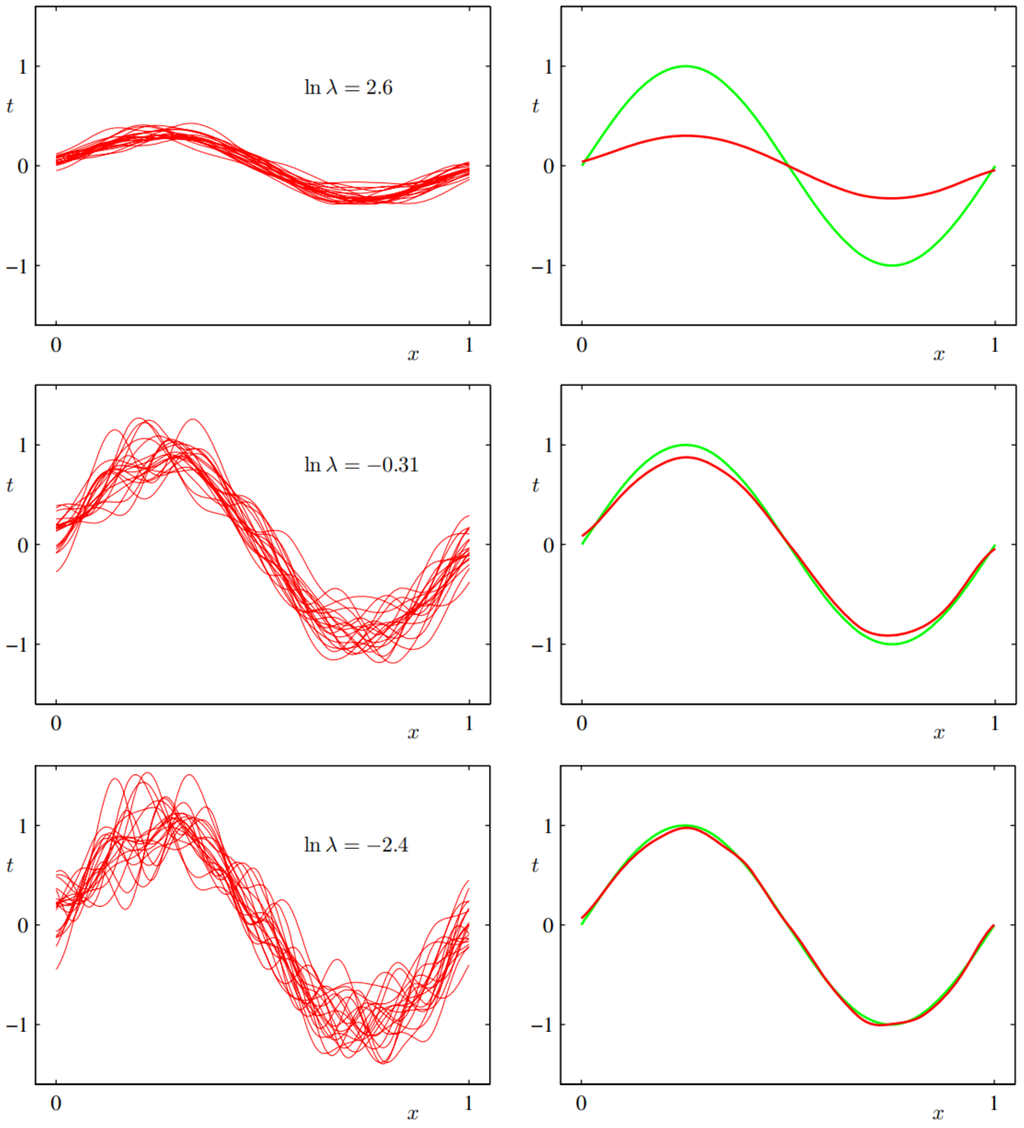
3.2 The Bias-Variance Decomposition

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}}\left[\{y(\mathbf{x};\mathcal{D}) - h(\mathbf{x})\}^2\right] \\
&\quad = \underbrace{\{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2}_{(\text{bias})^2} + \underbrace{\mathbb{E}_{\mathcal{D}}\left[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2\right]}_{\text{variance}}.
\end{aligned}
$$

$$
\text{expected loss} = (\text{bias})^2 + \text{variance} + \text{noise}
$$

$$
\begin{aligned}
(\text{bias})^2 &= \int \{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
\text{variance} &= \int \mathbb{E}_{\mathcal{D}}\left[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
\text{noise} &= \int \{h(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t
\end{aligned}
$$

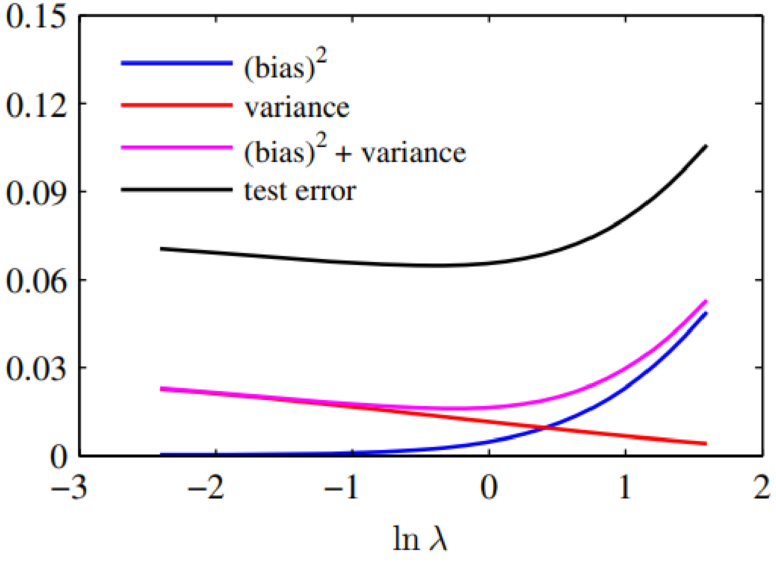
3.2 The Bias-Variance Decomposition



$$\text{expected loss} = (\text{bias})^2 + \text{variance} + \text{noise}$$

유연한 모델 : bias↓, variance↑

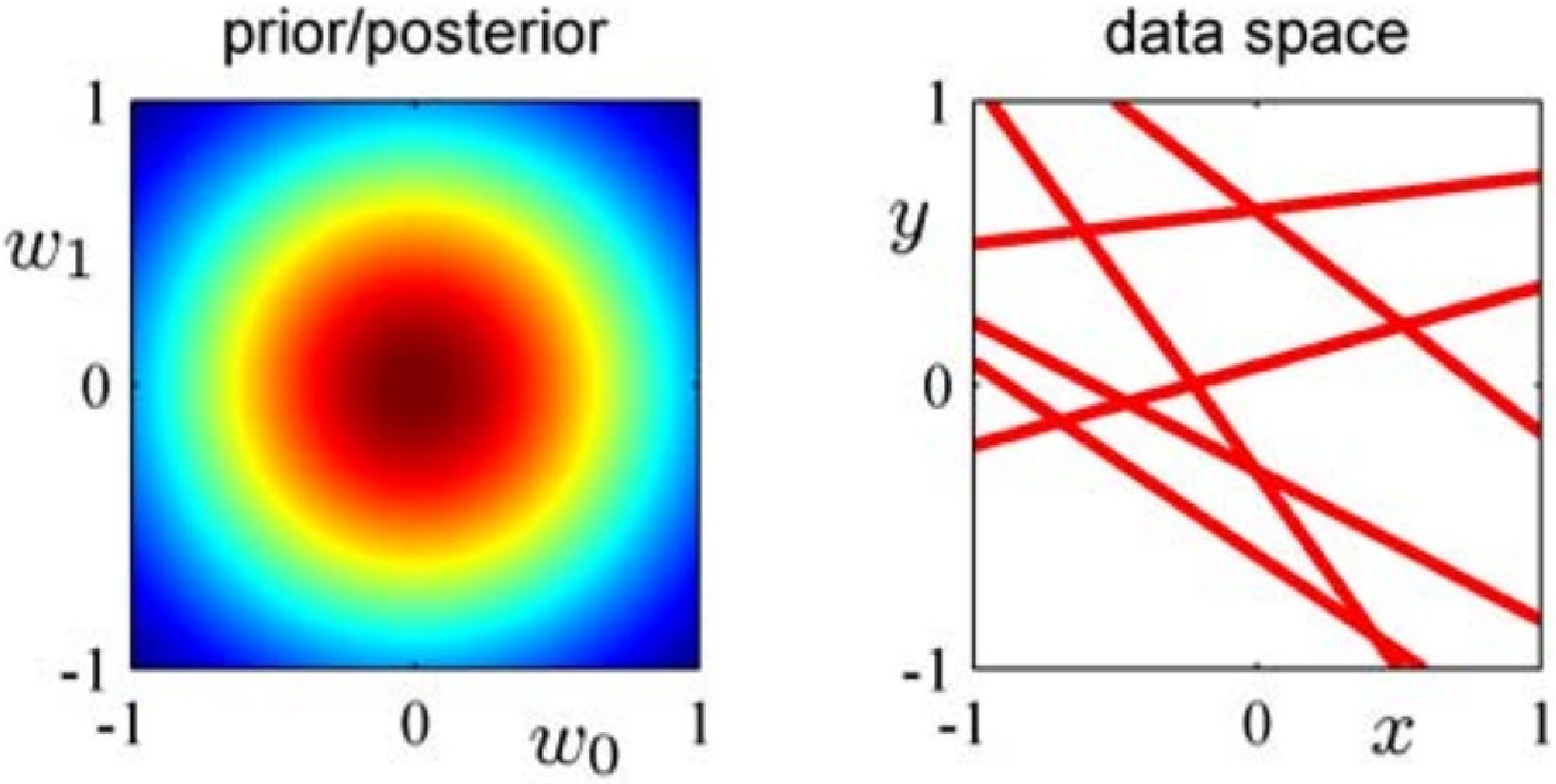
엄격한 모델 : bias ↑, variance ↓



3.3 Bayesian Linear Regression

3.3.1 parameter distribution

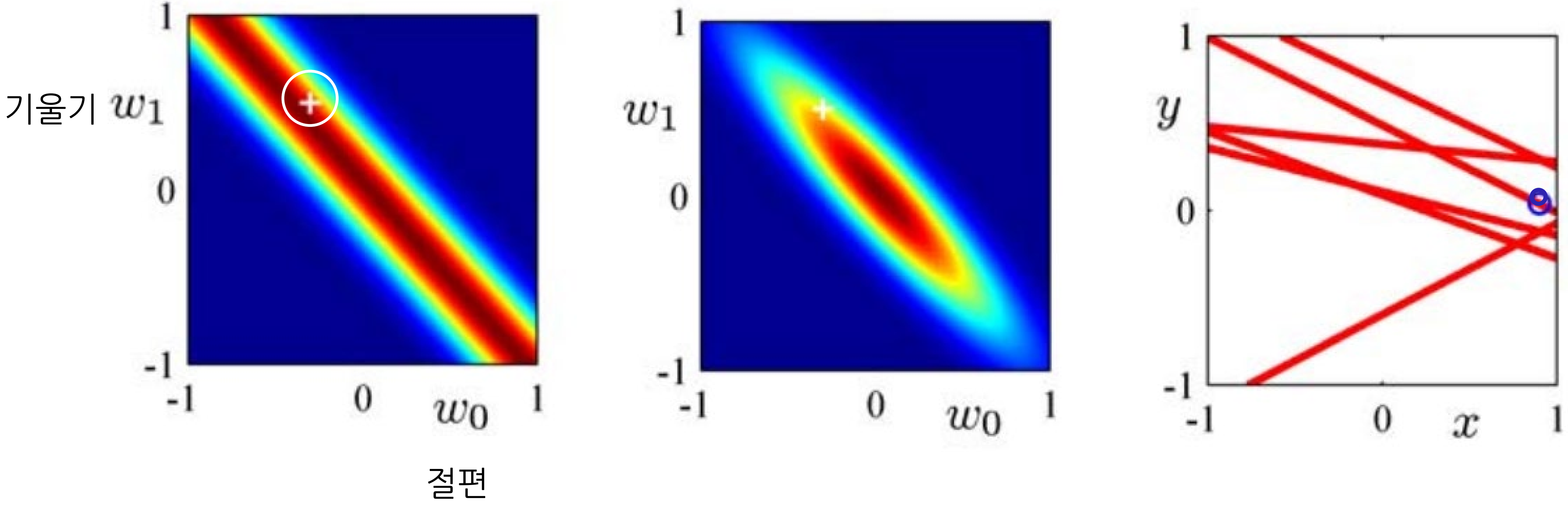
기울기



절편

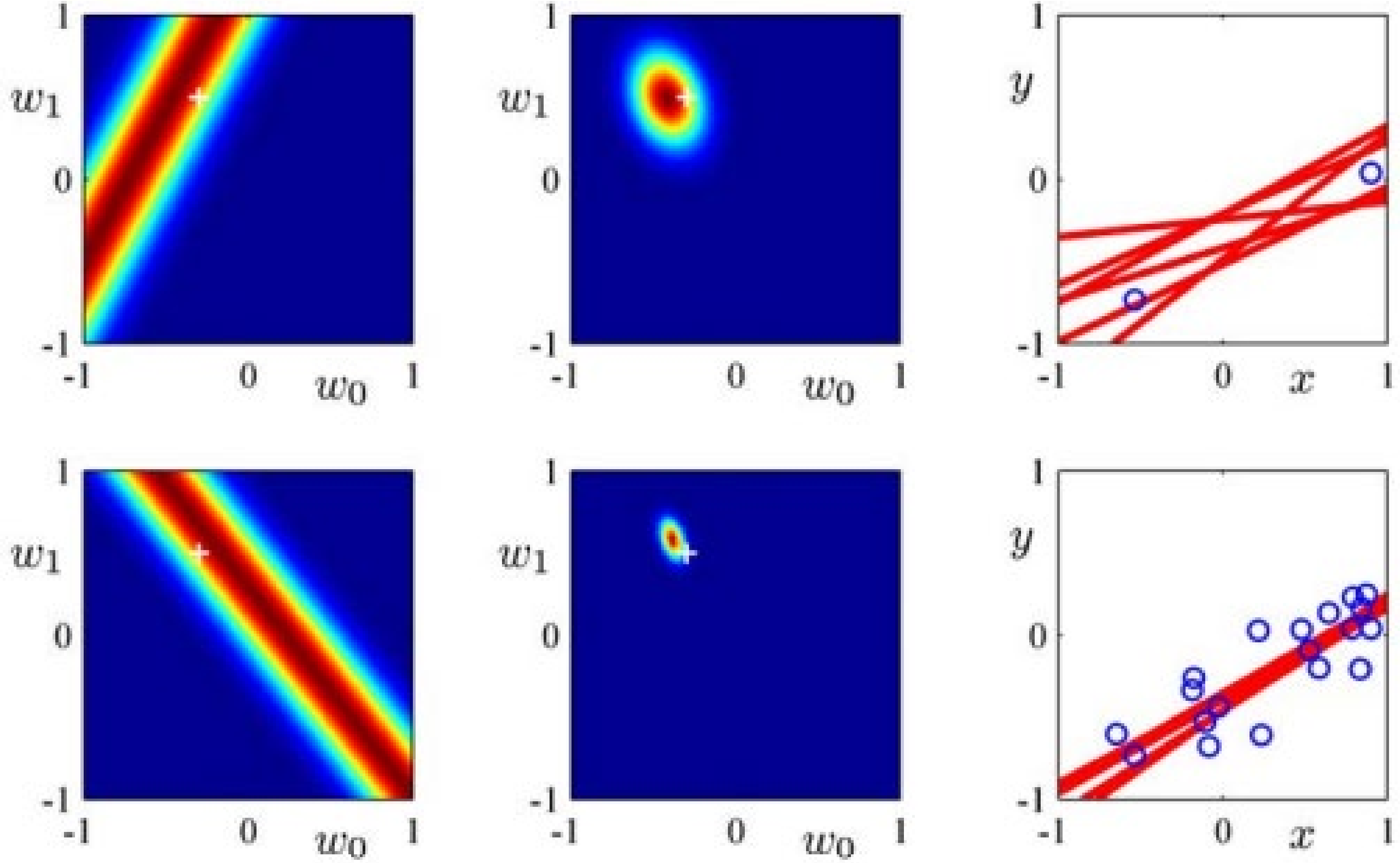
$$y(x, \mathbf{w}) = w_0 + w_1 x$$

3.3.1 parameter distribution



$$y(x, \mathbf{w}) = w_0 + w_1 x$$

3.3.1 parameter distribution



3.3.2 predictive distribution

사후 가중 분포

$$p(t|\mathbf{t},\alpha,\beta) = \int \underline{p(t|\mathbf{w},\beta)}\,\overline{p(\mathbf{w}|\mathbf{t},\alpha,\beta)}\,\mathrm{d}\mathbf{w}$$

두 가우시안 분포
의 콘볼루션

조건부 분포

기저함수

$$p(t|\mathbf{x},\mathbf{t},\alpha,\beta) = \mathcal{N}(t|\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}),\sigma_N^2(\mathbf{x}))$$

예측 분포

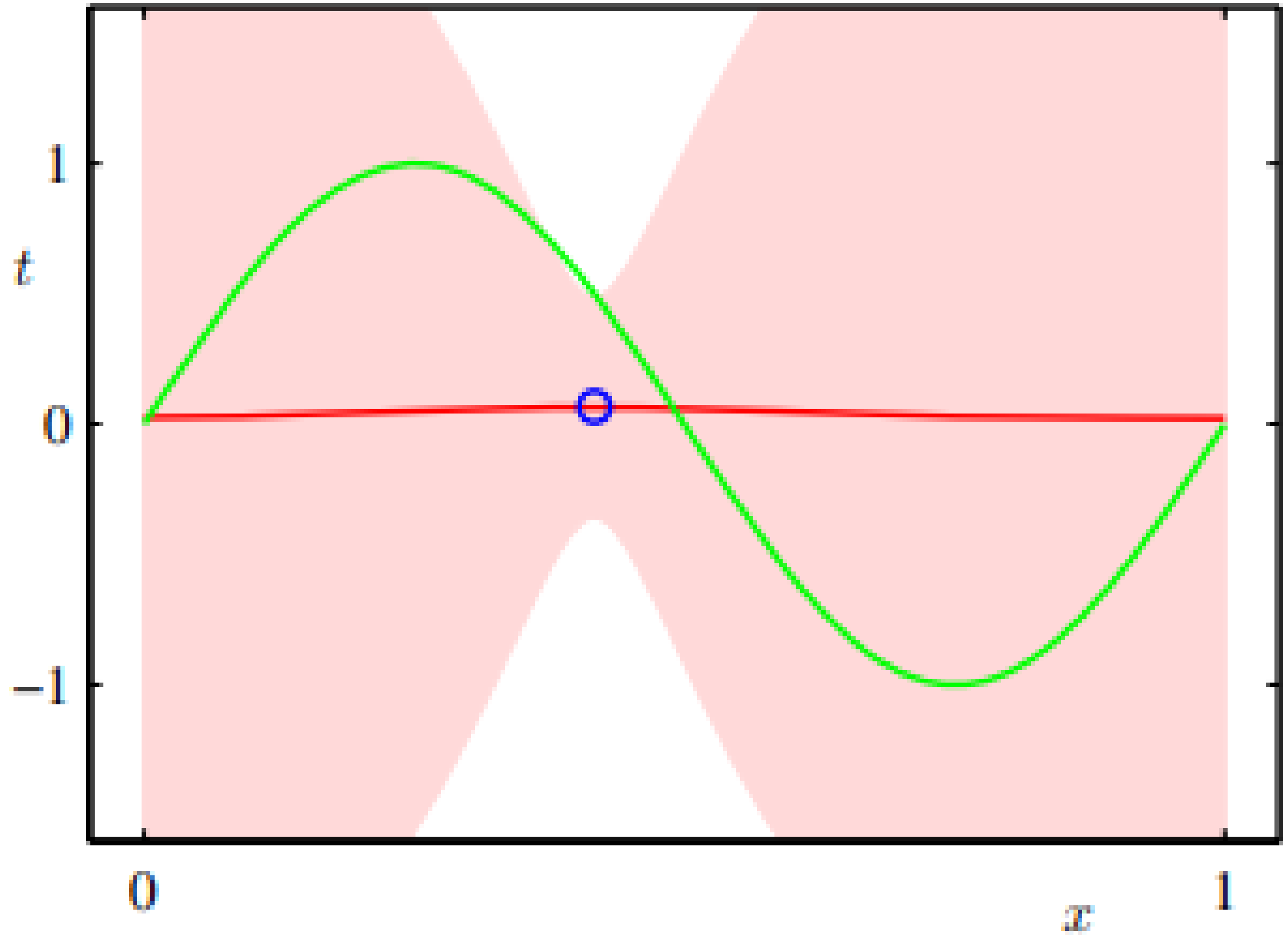
데이터 노이즈

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{S}_N\boldsymbol{\phi}(\mathbf{x}).$$

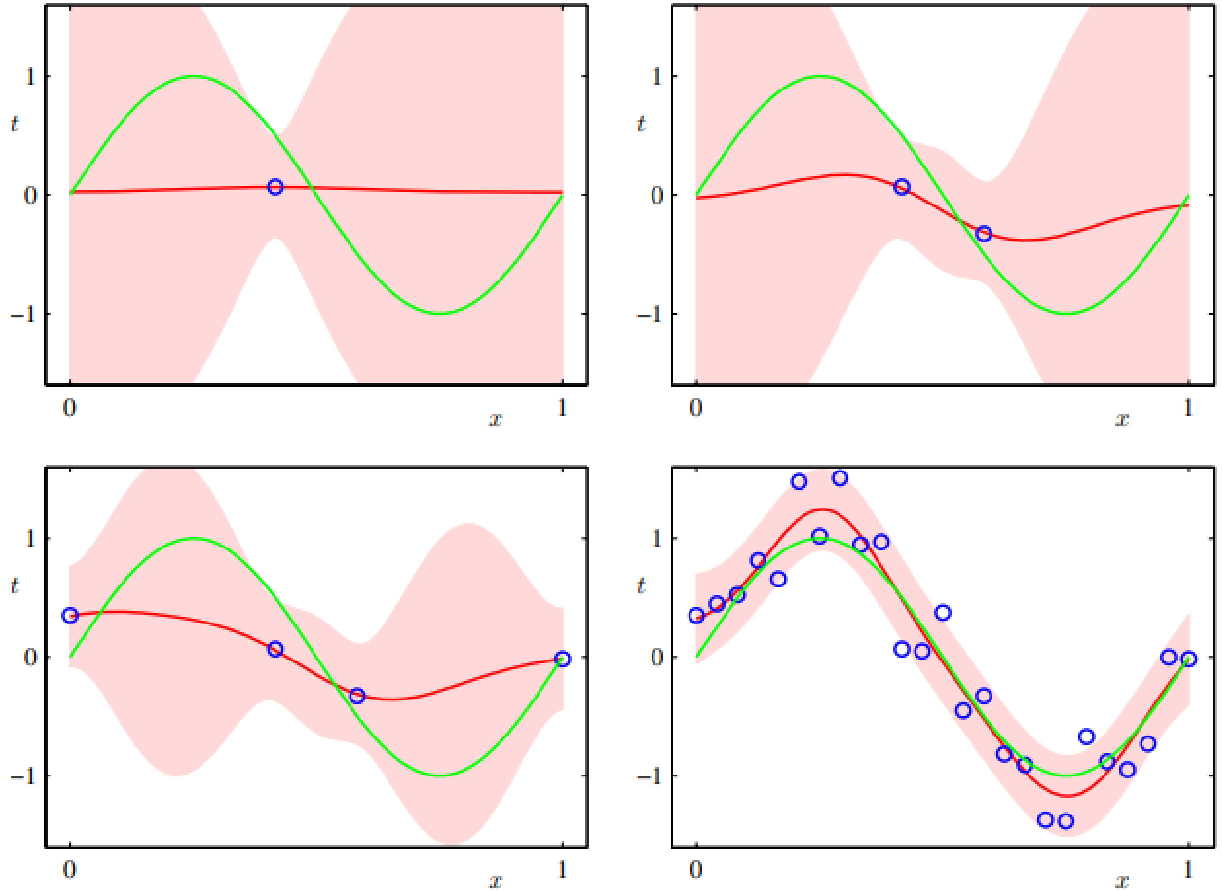
예측 분포의 분산

매개변수 w에 대한 불확실성

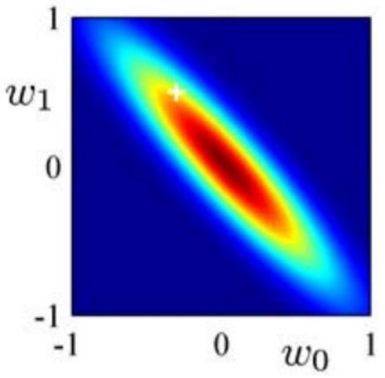
3.3.2 predictive distribution



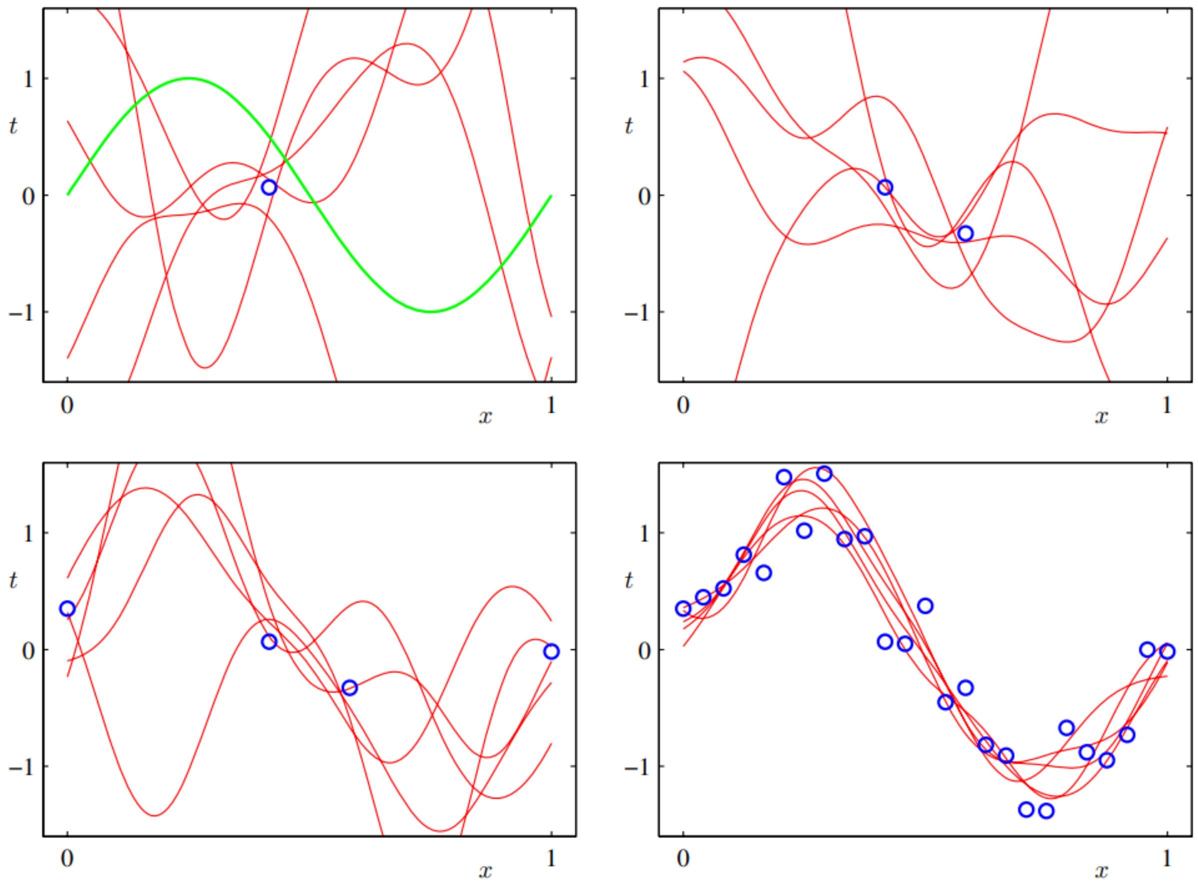
3.3.2 predictive distribution



3.3.2 predictive distribution



여기서 뽑은 건 아니지만 이런 사전/사후 분포에서 뽑아옴



3.3.3 equivalent kernel(smoother matrix)

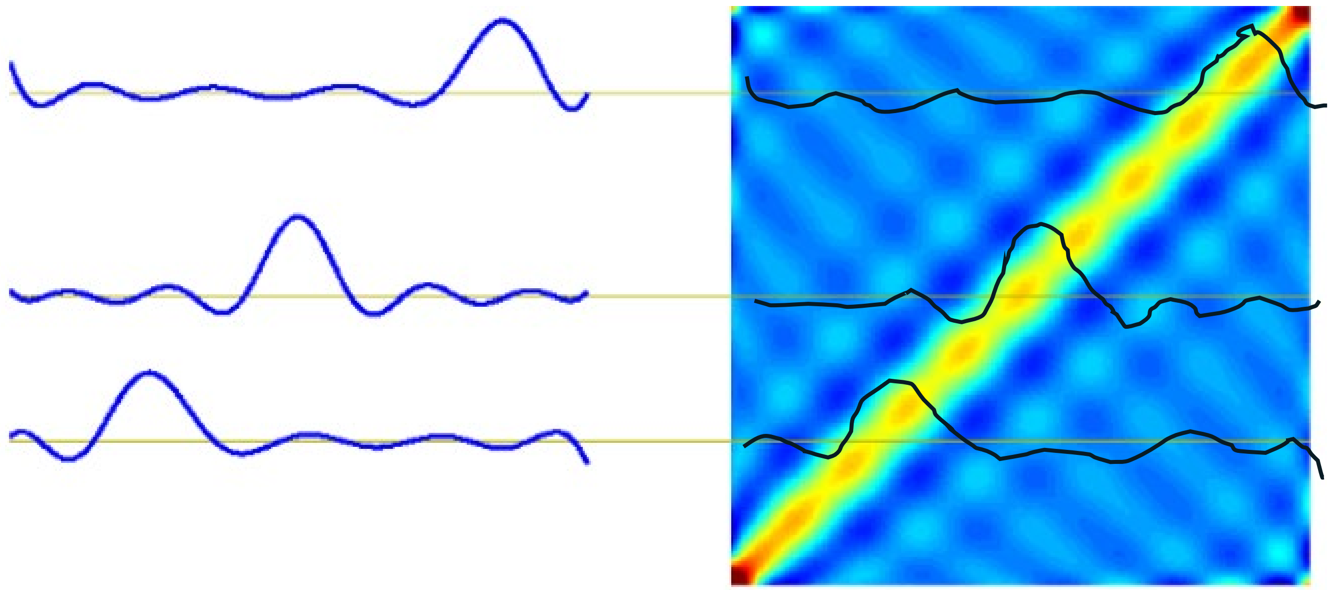
$$y(\mathbf{x}, \mathbf{m}_N) = \mathbf{m}_N^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}) = \beta \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} = \sum_{n=1}^{N} \beta \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}_n) t_n$$

$$y(\mathbf{x}, \mathbf{m}_N) = \sum_{n=1}^{N} k(\mathbf{x}, \mathbf{x}_n) t_n$$

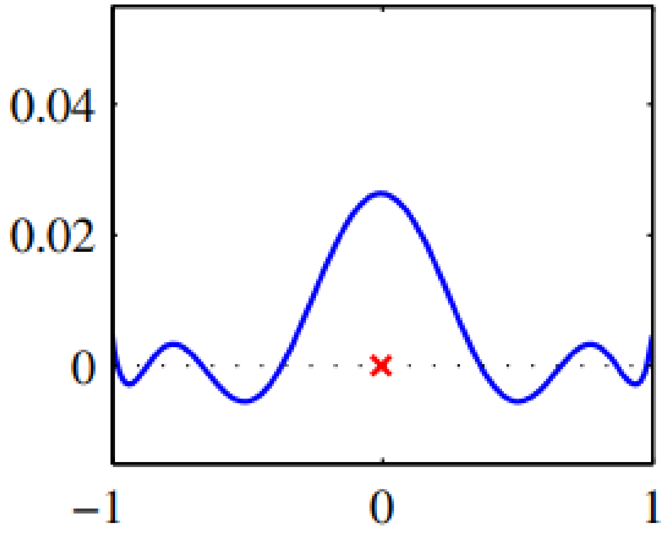
$$k(\mathbf{x}, \mathbf{x}') = \beta \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}')$$

3.3.3 equivalent kernel(smoother matrix)

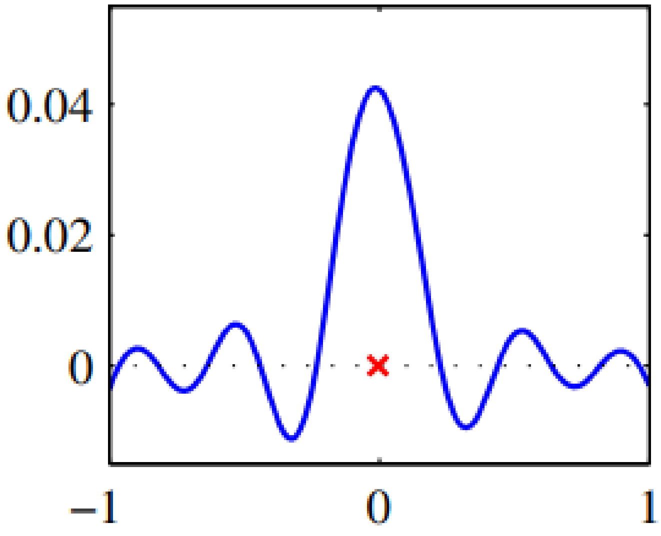
$$k(\mathbf{x}, \mathbf{x}') = \beta \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \boldsymbol{\phi}(\mathbf{x}')$$



3.3.3 equivalent kernel(smoother matrix)



다항 기저 함수



시그모이드 기저 함수

$$\begin{aligned}\mathrm{cov}[y(\mathbf{x}), y(\mathbf{x}')] &= \mathrm{cov}[\boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{w}, \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}')] \\ &= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{S}_N\boldsymbol{\phi}(\mathbf{x}') = \beta^{-1}k(\mathbf{x}, \mathbf{x}')\end{aligned}$$

3.3.3 equivalent kernel(smoother matrix)

$$k(\mathbf{x}, \mathbf{z}) = \psi(\mathbf{x})^{\mathrm{T}} \psi(\mathbf{z})$$

$$\text{where } \psi(\mathbf{x}) = \beta^{1/2} \mathbf{S}_N^{1/2} \phi(\mathbf{x})$$

3.4 Bayesian Model Comparison

3.4 Bayesian Model Comparison

불확실성 → 확률

모델 : 관측된 데이터 집합에 대한 확률 분포

$p(\mathcal{M}_i)$ L개의 모델

$$p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{M}_i)p(\mathcal{D}|\mathcal{M}_i).$$

모델 증거 = 주변 가능도

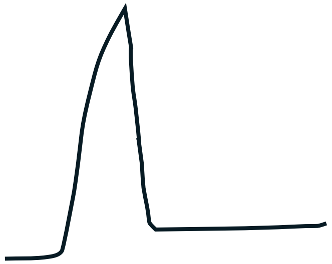
$p(\mathcal{D}|\mathcal{M}_i)/p(\mathcal{D}|\mathcal{M}_j)$ 베이즈 요인

3.4 Bayesian Model Comparison

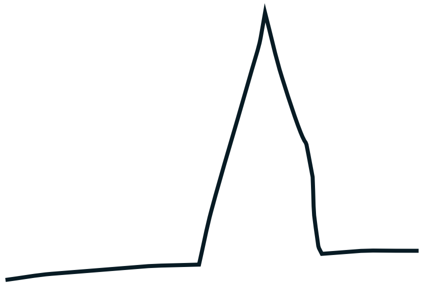
사후 확률

$$p(t|\mathbf{x},\mathcal{D}) = \sum_{i=1}^{L} p(t|\mathbf{x},\mathcal{M}_i,\mathcal{D})p(\mathcal{M}_i|\mathcal{D}).$$

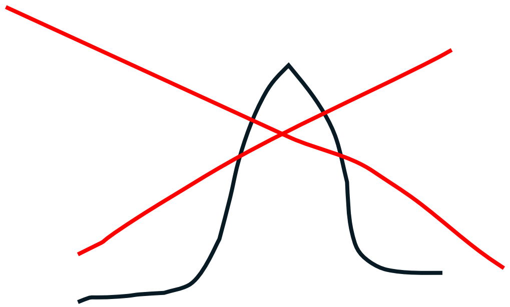
예측분포



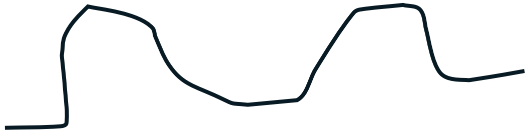
t=a



t=b



t=(a+b)/2



t=a t=b

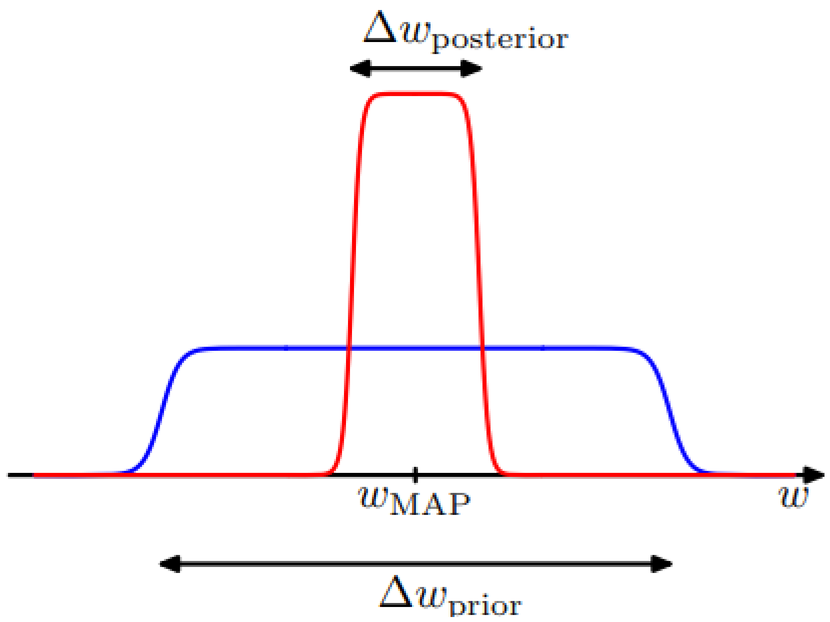
3.4 Bayesian Model Comparison

$$p(\mathcal{D}) = \int p(\mathcal{D}|w)p(w)\,\mathrm{d}w \simeq p(\mathcal{D}|w_{\mathrm{MAP}})\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}$$

$$\ln p(\mathcal{D}) \simeq \underline{\ln p(\mathcal{D}|w_{\mathrm{MAP}})} \overset{M}{+} \ln\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right)$$

가능성 높은 매개변수 바탕으로 근사한 것
(만약 평평한 사전 분포라면 로그 가능도)

모델 복잡도에 대한 penalize



3.4 Bayesian Model Comparison

$$p(\mathcal{D}) = \int p(\mathcal{D}|w)p(w)\,\mathrm{d}w \simeq p(\mathcal{D}|w_{\mathrm{MAP}})\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}$$

$$\ln p(\mathcal{D}) \simeq \underline{\ln p(\mathcal{D}|w_{\mathrm{MAP}})} + M \ln\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right)$$

가능성 높은 매개변수 바탕으로 근사한 것
(만약 평평한 사전 분포라면 로그 가능도)

모델 복잡도에 대한 penalize

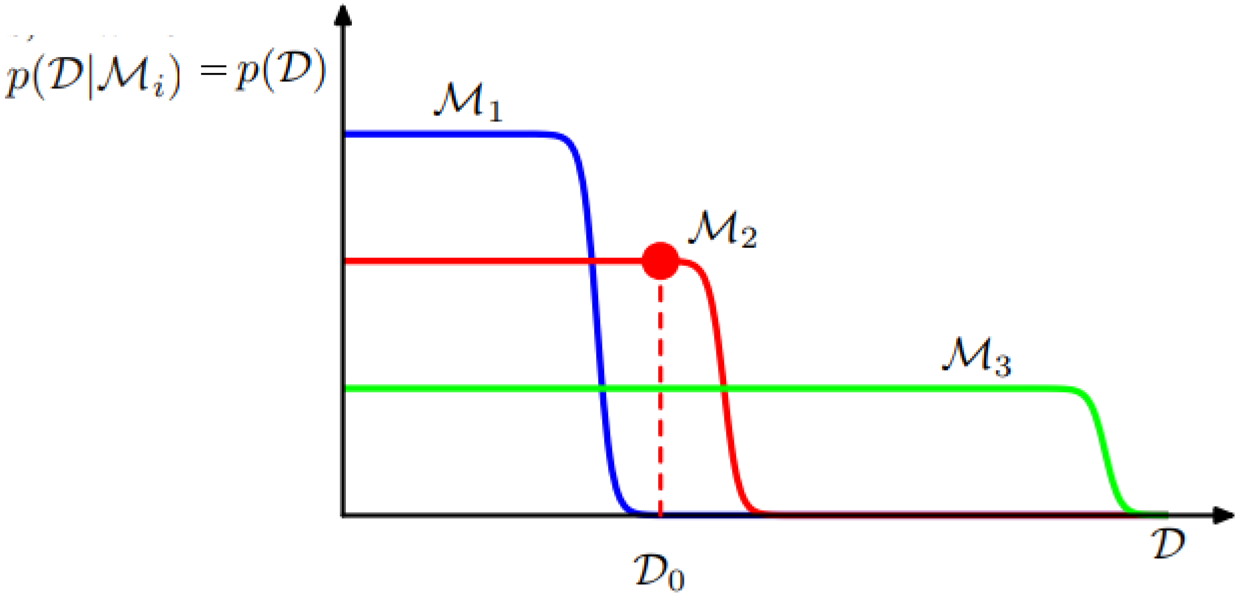
매개변수 선택

$p(\mathbf{w})$

$\downarrow$

$p(\mathcal{D}|\mathbf{w})$

데이터 추출



3.4 Bayesian Model Comparison

$$p(\mathcal{D}) = \int p(\mathcal{D}|w)p(w)\,\mathrm{d}w \simeq p(\mathcal{D}|w_{\mathrm{MAP}})\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}$$

$$\ln p(\mathcal{D}) \simeq \underline{\ln p(\mathcal{D}|w_{\mathrm{MAP}})} + M \ln\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right)$$

가능성 높은 매개변수 바탕으로 근사한 것
(만약 평평한 사전 분포라면 로그 가능도)

모델 복잡도에 대한 penalize

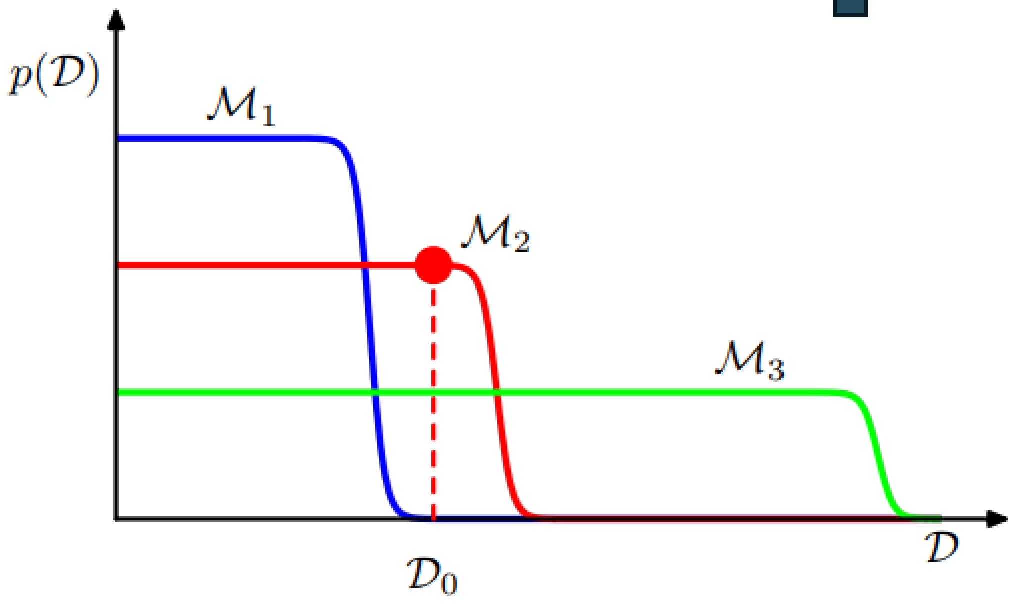
매개변수 선택

$p(\mathbf{w})$

$\downarrow$

$p(\mathcal{D}|\mathbf{w})$

데이터 추출



$$\int p(\mathcal{D}|\mathcal{M}_1)\ln\frac{p(\mathcal{D}|\mathcal{M}_1)}{p(\mathcal{D}|\mathcal{M}_2)}\,\mathrm{d}\mathcal{D}$$

3.5 The Evidence Approximation

$$p(t|\mathbf{t}) = \iiint p(t|\mathbf{w},\beta)p(\mathbf{w}|\mathbf{t},\alpha,\beta)\underbrace{p(\alpha,\beta|\mathbf{t})}_{\text{사후 분포}}\,\mathrm{d}\mathbf{w}\,\mathrm{d}\alpha\,\mathrm{d}\beta$$

$$p(t|\mathbf{t}) \simeq p(t|\mathbf{t},\widehat{\alpha},\widehat{\beta}) = \int p(t|\mathbf{w},\widehat{\beta})p(\mathbf{w}|\mathbf{t},\widehat{\alpha},\widehat{\beta})\,\mathrm{d}\mathbf{w}.$$

$$p(\alpha,\beta|\mathbf{t}) \propto p(\mathbf{t}|\alpha,\beta)p(\alpha,\beta)$$

3.5.1Evaluation of the evidence function

가중 매개변수 w에 대해서 적분

$$p(\mathbf{t}|\alpha, \beta) = \int p(\mathbf{t}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)\,\mathrm{d}\mathbf{w}.$$

주변 가능 도함수

$$p(\mathbf{t}|\alpha, \beta) = \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\{-E(\mathbf{w})\}\,\mathrm{d}\mathbf{w}$$

$$\begin{aligned} E(\mathbf{w}) &= \beta E_D(\mathbf{w}) + \alpha E_W(\mathbf{w}) \\ &= \frac{\beta}{2}\|\mathbf{t} - \mathbf{\Phi}\mathbf{w}\|^2 + \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}. \end{aligned}$$

$$E(\mathbf{w}) = E(\mathbf{m}_N) + \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N)$$

M은 w의 차원수

3.5.1 Evaluation of the evidence function

$$\mathbf{A} = \nabla\nabla E(\mathbf{w})$$

$$\mathbf{m}_N = \beta \mathbf{A}^{-1} \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t}.$$

$$\mathbf{S}_N^{-1} = \alpha \mathbf{I} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}.$$

$$\mathbf{A} = \alpha \mathbf{I} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}$$

$$E(\mathbf{m}_N) = \frac{\beta}{2} \|\mathbf{t} - \boldsymbol{\Phi}\mathbf{m}_N\|^2 + \frac{\alpha}{2} \mathbf{m}_N^{\mathrm{T}} \mathbf{m}_N.$$

$$\begin{aligned} & \int \exp\{-E(\mathbf{w})\}\, \mathrm{d}\mathbf{w} \\ &= \exp\{-E(\mathbf{m}_N)\} \int \exp\left\{-\frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}} \mathbf{A} (\mathbf{w} - \mathbf{m}_N)\right\} \mathrm{d}\mathbf{w} \\ &= \exp\{-E(\mathbf{m}_N)\} (2\pi)^{M/2} |\mathbf{A}|^{-1/2}. \end{aligned}$$

가중 매개변수 w에 대해서 적분

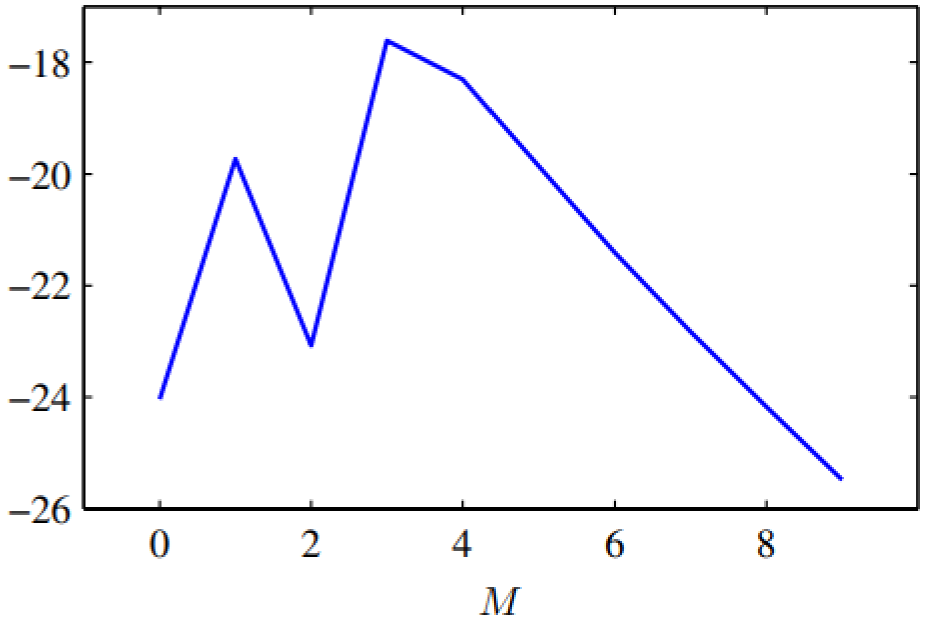
3.5.1 Evaluation of the evidence function

$$\ln p(\mathbf{t}|\alpha,\beta) = \frac{M}{2}\ln\alpha + \frac{N}{2}\ln\beta - E(\mathbf{m}_N) - \frac{1}{2}\ln|\mathbf{A}| - \frac{N}{2}\ln(2\pi)$$

이용!

$$p(\mathbf{t}|\alpha,\beta) = \int p(\mathbf{t}|\mathbf{w},\beta)p(\mathbf{w}|\alpha)\,\mathrm{d}\mathbf{w}.$$

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0},\alpha^{-1}\mathbf{I}) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2}\exp\left\{-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}$$



3.5.2 Maximizing the evidence function

가중 매개변수 w에 대해서 적분

$$p(\mathbf{t}|\alpha, \beta) = \int p(\mathbf{t}|\mathbf{w}, \beta) p(\mathbf{w}|\alpha)\, \mathrm{d}\mathbf{w}.$$

$$\left(\beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}\right) \mathbf{u}_i = \lambda_i \mathbf{u}_i.$$

$$\mathbf{A} = \alpha \mathbf{I} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}$$

α에 대해서 미분

$$\ln p(\mathbf{t}|\alpha, \beta) = \frac{M}{2} \ln \alpha + \frac{N}{2} \ln \beta - E(\mathbf{m}_N) - \frac{1}{2} \ln |\mathbf{A}| - \frac{N}{2} \ln(2\pi)$$

3.5.2 Maximizing the evidence function

$$\alpha \mathbf{m}_N^{\mathrm{T}} \mathbf{m}_N = M - \alpha \sum_i \frac{1}{\lambda_i + \alpha} = \gamma.$$

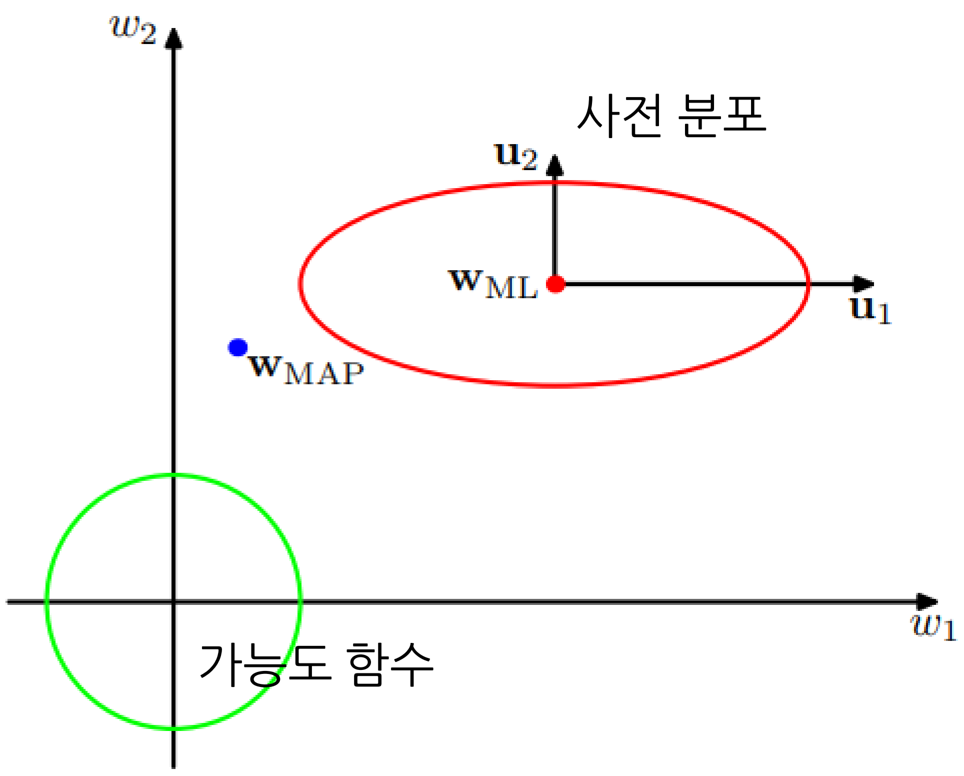
$$\gamma = \sum_i \frac{\lambda_i}{\alpha + \lambda_i}.$$

$$\mathbf{m}_N = \beta \mathbf{A}^{-1} \mathbf{\Phi}^{\mathrm{T}} \mathbf{t}.$$

$$\alpha = \frac{\gamma}{\mathbf{m}_N^{\mathrm{T}} \mathbf{m}_N}.$$

α에 대해서 미분

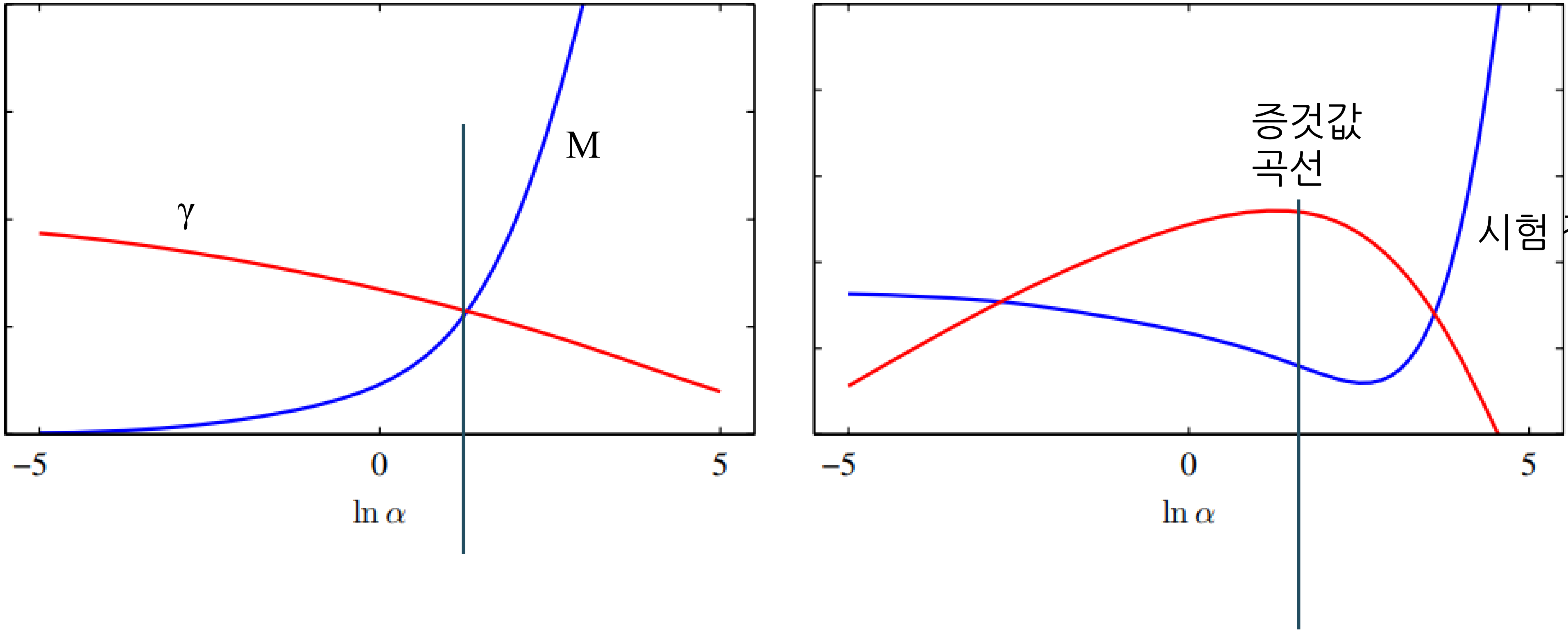
3.5.3 Effective number of parameters



$$\mathbf{m}_N = \beta \mathbf{A}^{-1} \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t}.$$

$$\left(\beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}\right) \mathbf{u}_i = \lambda_i \mathbf{u}_i .$$

3.5.3 Effective number of parameters



$$\alpha = \frac{M}{2E_W(\mathbf{m}_N)}$$

$$\beta = \frac{N}{2E_D(\mathbf{m}_N)}$$

이 결과

3.6 Limitations of Fixed Basis Functions

매개변수 선형성 가정
→ 제곱 문제의 해가 닫힌 형태로 존재, 베이지안 과정 통해 풀이 가능

기저 함수 적절히 선택했다면

→ x와 t사이의 비선형성도 모델 가능

3.6 Limitations of Fixed Basis Functions

선형 모델의 한계점
(해결 : 서포트 벡터 머신, 뉴럴 네트워크)

기저 함수가 훈련 데이터 집합 관측 전에 고정
→ 차원의 저주 문제
→ 기저 함수의 숫자는 입력 공간의 차원이 증가함에 따라 빠르게 증가

3.6 Limitations of Fixed Basis Functions

실제 데이터 집합들의 두 성질 (1)

데이터 벡터들은 보통 내재적 차원수가 입력 공간의 차원수보다 작은 비선형 매니폴드에 근접 가능
→ 입력 변수 사이엔 강한 상관관계 존재

방사 기저 함수 네트워크, 서포트 벡터 머신, 상관 벡터 머신 등에서 사용 가능

3.6 Limitations of Fixed Basis Functions

실제 데이터 집합들의 두 성질 (2)

타깃 변수들이 데이터 매니폴드의 몇몇 일부 방향성에 대해서만 중요한 종속성을 가졌을 수 있음

Thank you