10.6 変分ロジスティック回帰

平成 28 年 9 月 11 日

概 要

PRML の「10.6 変分ロジスティック回帰」についての実装と考察

目 次

1　問題設定

ロジスティック回帰で超パラメータの推定を行うときには近似が必要となる. ここでは変分近似を用いる.

2　10.6.1, 10.6.2 変分ロジスティック回帰

2.1　アルゴリズム

変分法の枠組みは, 尤度関数の下界を最大化することが目標である.
ベイズロジスティック回帰モデルの尤度関数は

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{w})p(\mathbf{w})d\mathbf{w} = \int \left[\prod_{n=1}^{N} p(t_n|\mathbf{w})\right] p(\mathbf{w})d\mathbf{w} \quad (10.147)$$

ただし, t の条件付き分布は $a = \mathbf{w}^T\boldsymbol{\phi}$ として

$$p(t|\mathbf{w}) = \sigma(a)^t\{1-\sigma(a)\}^{1-t} = e^{at}\sigma(-a) \quad (10.148)$$

となっている.
$p(\mathbf{t})$ の下界を得るために, まずロジスティックシグモイド関数の変分下界

$$\sigma(z) \geq \sigma(\xi)exp\{(z-\xi)/2 - \lambda(\xi)(z^2-\xi^2)\} \quad (10.149), \quad \lambda(\xi) = \frac{1}{2\xi}\left[\sigma(\xi) - \frac{1}{2}\right] \quad (10.150)$$

を利用し,

$$p(t|\mathbf{w}) \geq e^{at}\sigma(\xi)exp\{-(a+\xi)/2 - \lambda(\xi)(a^2-\xi^2)\} \quad (10.151)$$

とすることができる.
以上より, $\mathbf{t}$ と $\mathbf{w}$ の同時分布は

$$p(\mathbf{t},\mathbf{w}) = p(\mathbf{t}|\mathbf{w})p(\mathbf{w}) \geq h(\mathbf{w},\boldsymbol{\xi})p(\mathbf{w}) \quad (10.152)$$

ここで $h(\mathbf{w},\boldsymbol{\xi})$ は

$$h(\mathbf{w},\boldsymbol{\xi}) = \prod_{n=1}^{N} \sigma(\xi_n)exp\{\mathbf{w}^T\boldsymbol{\phi}_n t_n - (\mathbf{w}^T\boldsymbol{\phi}_n + \xi_n)/2 - \lambda(\xi_n)([\mathbf{w}^T\boldsymbol{\phi}_n]^2 - \xi_n^2)\} \quad (10.153)$$

これより, 対数同時分布は, $\mathbf{w}$ に依存する項でまとめると

$$\begin{aligned}\ln\{p(\boldsymbol{t}|\mathbf{w})p(\mathbf{w})\} &= \ln p(\mathbf{w}) + \sum_{n=1}^{N}\{\ln\sigma(\xi_n) + \mathbf{w}^T\boldsymbol{\phi}_n t_n - (\mathbf{w}^T\boldsymbol{\phi}_n + \xi_n)/2 - \lambda(\xi_n)([\mathbf{w}^T\boldsymbol{\phi}_n]^2 - \xi_n^2) \\ &= -\frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^T S_0^{-1}(\mathbf{w}-\mathbf{m}_0) + \sum_{n=1}^{N}\{\mathbf{w}^T\boldsymbol{\phi}_n(t_n - 1/2) - \lambda(\xi_n)\mathbf{w}^T(\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T)\mathbf{w}\} + const\end{aligned}$$

となる, これは W の変分事後分布がガウス分布であることを示す.

$$q(\mathbf{w}) = N(\mathbf{w}|\mathbf{m}_N, S_N) \quad (10.156)$$

ここで,

$$\mathbf{m}_N = S_N\left(S_0^{-1}\mathbf{m}_0 + \sum_{n=1}^{N}(t_n - 1/2)\boldsymbol{\phi}_n\right) \quad (10.157), \quad S_N^{-1} = S_0^{-1} + 2\sum_{n=1}^{N}\lambda(\xi_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \quad (10.158)$$

である.
次は, 変分パラメータ $\boldsymbol{\xi}$ の最適化を考える. ここでは, EM アルゴリズムを用いる.
まず, E ステップではパラメータ $\boldsymbol{\xi}^{old}$ を用いて W の事後分布 (10.156) を求める.
次に, M ステップでは次式の期待完全データ対数尤度を最大化する.

$$\begin{aligned} Q(\boldsymbol{\xi}, \boldsymbol{\xi}^{old}) &= E[\ln\{h(\mathbf{w}, \boldsymbol{\xi})p(W)\}] \quad (10.160) \\ &= \sum_{n=1}^{N}\{\ln\sigma(\xi_n) - \xi_n/2 - \lambda(\xi_n)(\boldsymbol{\phi}_n^T E[\mathbf{w}\mathbf{w}^T]\boldsymbol{\phi}_n - \xi_n^2)\} + const \quad (10.161) \end{aligned}$$

ξ_n についての微分を 0 とすることで

$$(\xi^{new})^2 = \boldsymbol{\phi}_n^T E[\mathbf{w}\mathbf{w}^T]\boldsymbol{\phi}_n = \boldsymbol{\phi}_n^T(S_N + \mathbf{m}_N\mathbf{m}_N^T)\boldsymbol{\phi}_n \quad (10.163)$$

を得る. この, E,M ステップを繰り返す.

変分ロジスティック回帰

1. まずパラメータ $\boldsymbol{\xi}$ を初期化する.
2. (E ステップ)$q(\mathbf{w})$ のパラメータを推定する.
3. (M ステップ)$\boldsymbol{\xi}$ を推定する.
4. 2,3 を繰り返す.

2.2 コード

変分ロジスティック回帰のコード (variational_logistic_regression.py).

```
for N in [10,50,100,200]:
  x=data[:N,:2]
  #t=1 or 0
  t=data[:N,2]

  #Phi=(1,x1,x2)
  M=3
  P=np.ones((N,M))
  P[:,1:3]=x

  xi=rd.rand(N)
  lam=np.zeros(N)
  m0=np.zeros(M)
  S0=np.identity(M)
  mN=np.zeros(M)
  SN=np.identity(M)

  diff=10**6
  while diff>=10**-6:
    diff=0
    for n in range(N):
      lam[n]=(sig(xi[n])-1/2)/(2*xi[n])

    tmp2=SN
    tmp1=np.zeros((M,M))
    for n in range(N):
      tmp1+=lam[n]*outer(P[n,:],P[n,:])
    tmp1*=2
```

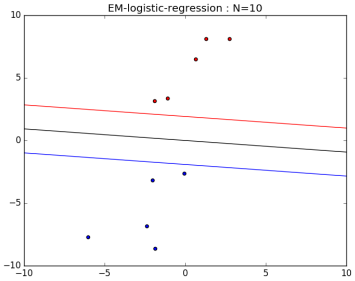
```
    SN=inv(inv(S0)+tmp1)
    diff+=norm(tmp2-SN)

    tmp2=mN
    tmp1=dot(inv(S0),m0)
    for n in range(N):
      tmp1+=(t[n]-1/2)*P[n,:]
    mN=dot(SN,tmp1)
    diff+=norm(tmp2-mN)

    tmp2=xi
    for n in range(N):
      xi[n]=sqrt(dot(P[n,:],dot(SN+outer(mN,mN),P[n,:])))
    diff+=norm(tmp2-xi)

  def model(z):
    y=np.array([1,z[0],z[1]])
    return sig(dot(mN,y))
```

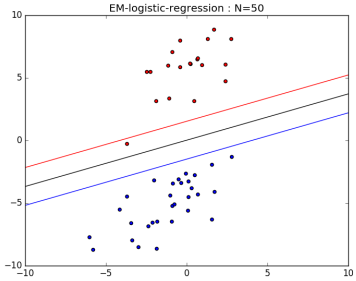
2.3 結果

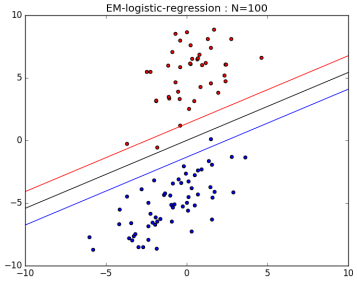
$N = 10, 50, 100$ に対して (内積基底以外)$M = 4, 10, 20$ として分類を行った.
以下では, 黒い線が決定境界, 赤い線が赤い点の確率 0.9, 青い線が青い点の確率 0.9 の線である.

2.3.1 内積基底

基本としてまず, 線形分離を行うために $\phi = (1, x_1, x_2)^T$ を用いた.

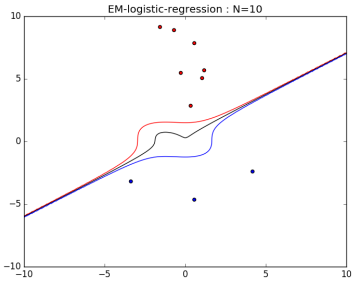


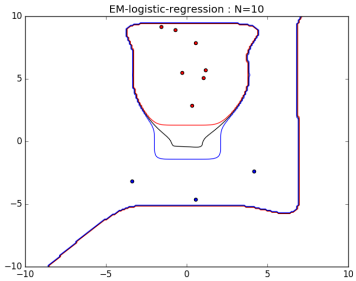


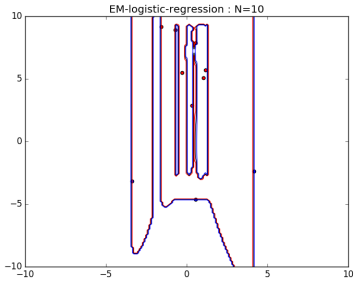


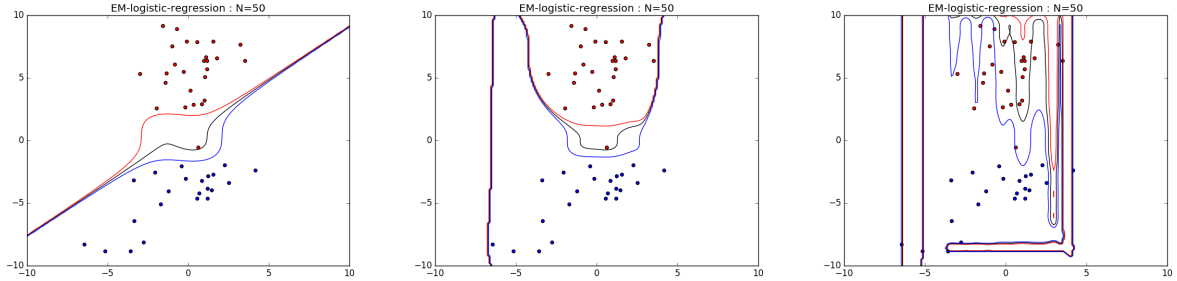
2.3.2 多項式基底

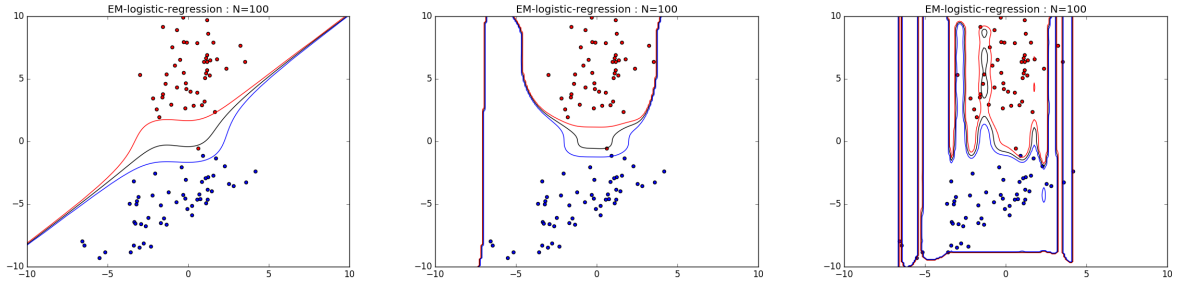
x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = x^i$ を選んだ.







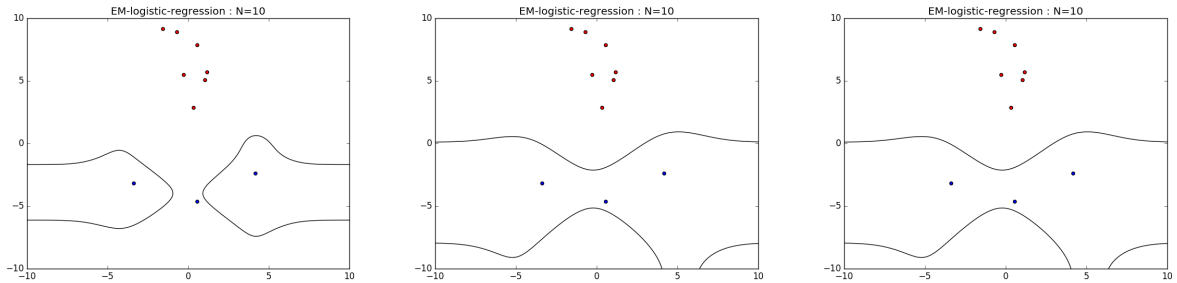


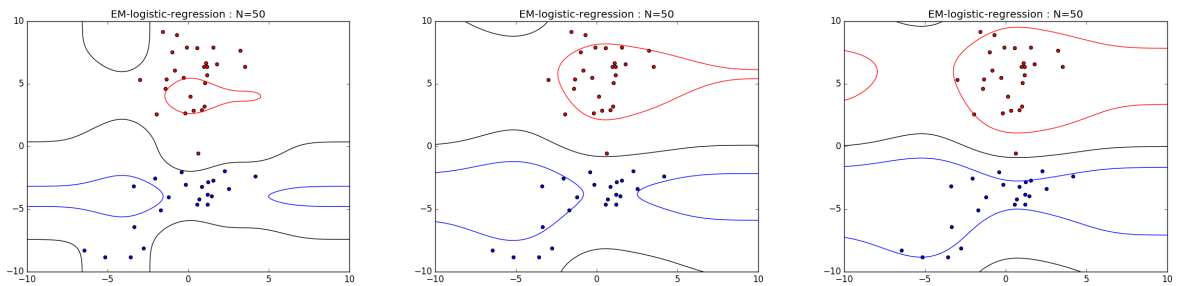


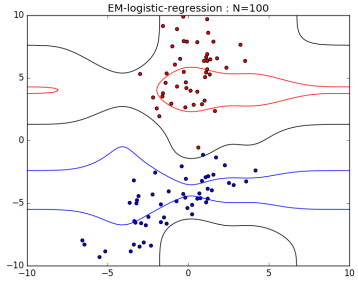
2.3.3 ガウス基底

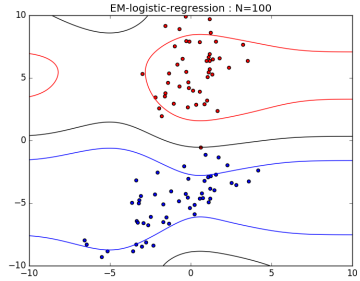
x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = exp\{-\frac{(x-\mu_i)^2}{2s^2}\}$ をもちいる．またここでは μ は区間 $[-10, 10]$ を M 個に等分する区間の中心を用い,s には 1.5 を用いた.

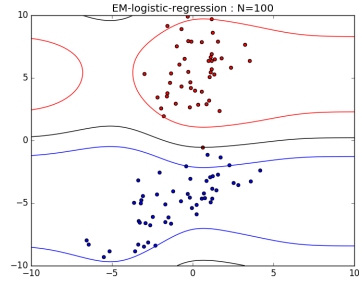
$$mu = [(m + 0.5) * (20/(M + 1)) - 10 \ for \ m \ in \ range(M)] \quad s = 1.5$$





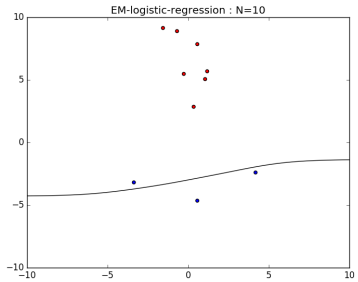


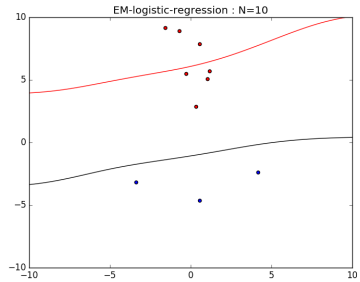


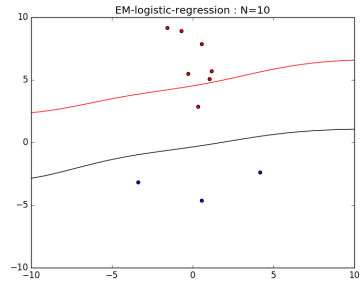


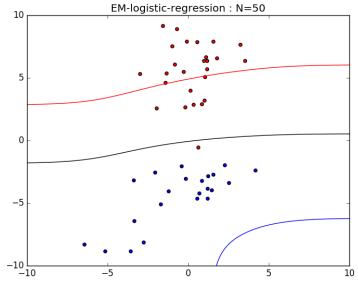
2.3.4 シグモイド基底

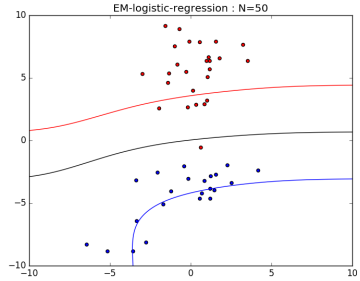
x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = \sigma(\frac{x-\mu_i}{s})$ をもちいる. ただし,$\sigma(a) = \frac{1}{1+e^{-a}}$

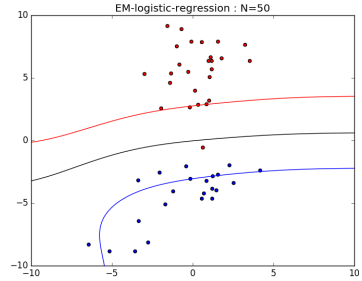


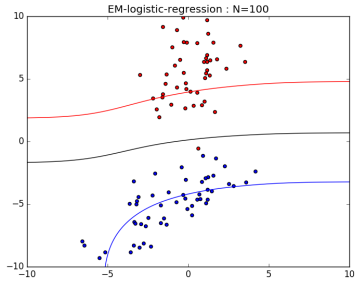


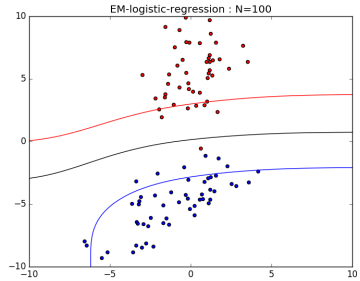


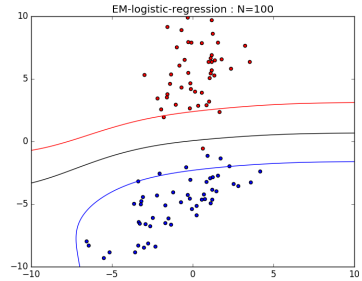






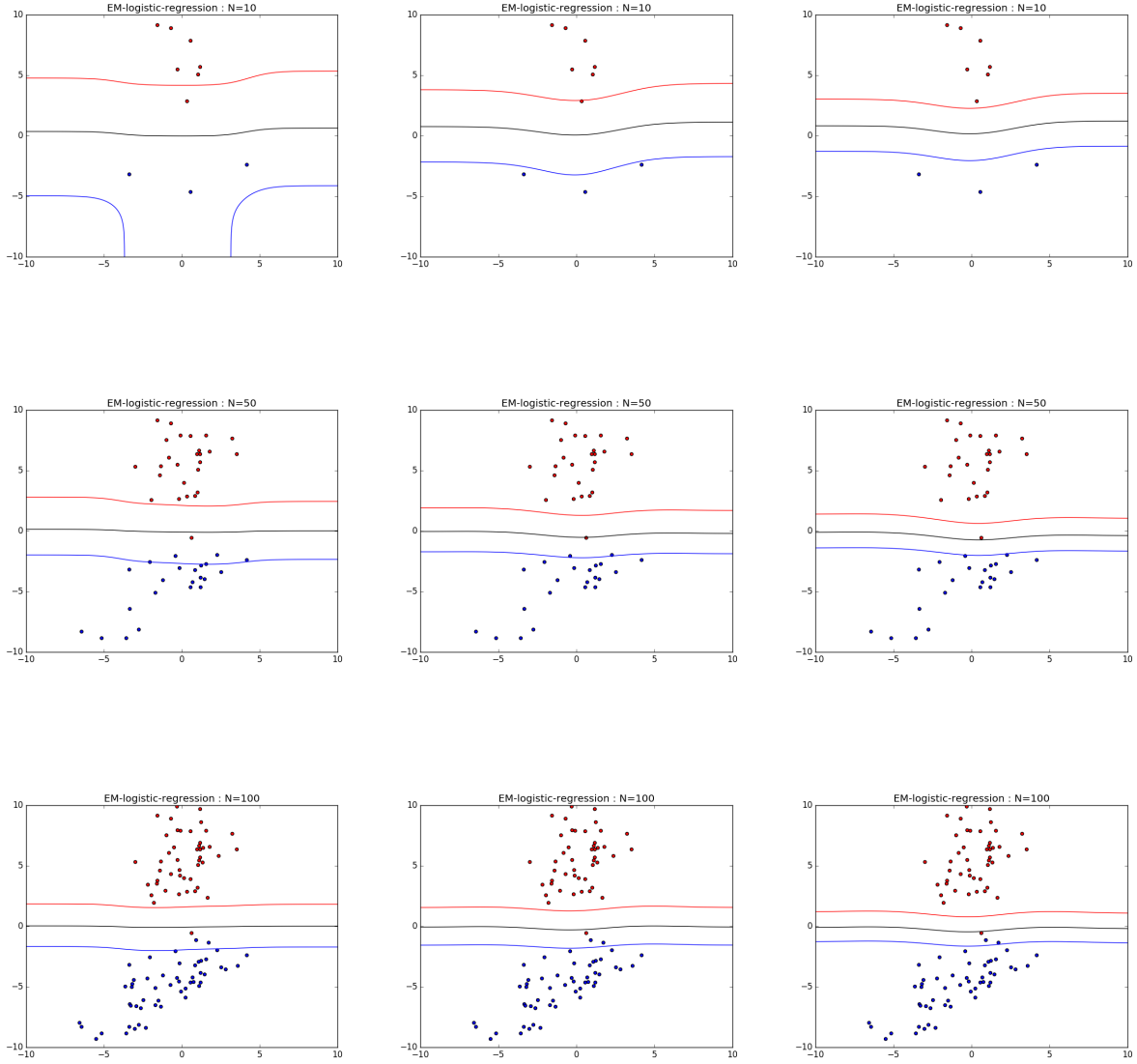






2.3.5 tanh 基底

x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = tanh(\frac{x-\mu_i}{s})$ をもちいる.



3　10.6.3 超パラメータの推定

ここまでは, 超パラメータについては考えず, 変分推定というよりは EM アルゴリズムって感じであった. 以降では, 超パラメータについて考え, 変分推定ぽくなる.

3.1　アルゴリズム

ここでは, 次の単純な等方ガウス分布

$$p(\mathbf{w}|\alpha) = N(\mathbf{w}|\mathbf{0}, \alpha^{-1}I) \quad (10.165)$$

とし, α に対する共役事前分布をガンマ分布

$$p(\alpha) = Gam(\alpha|a_0, b_0) \quad (10.166)$$

とする.
このモデルの周辺分布は

$$p(\mathbf{t}) = \int \int p(\mathbf{w}, \alpha, \mathbf{t}) d\mathbf{w} d\alpha \quad (10.167)$$

また, 同時分布は

$$p(\mathbf{w}, \alpha, \mathbf{t}) = p(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\alpha)p(\alpha) \quad (10.168)$$

で与えられる.
まず, 変分分布 $q(\mathbf{w}, \alpha)$ を導入し,

$$\ln p(\boldsymbol{t}) = L(q) + KL(q||p) \quad (10.169)$$

と分解でき

$$L(q) = \int\int q(W, \alpha) \ln \left\{ \frac{p(\mathbf{w}, \alpha, \mathbf{t})}{q(\mathbf{w}, \alpha)} \right\} d\mathbf{w} d\alpha \quad (10.170)$$

$$KL(q||p) = -\int\int q(\mathbf{w}, \alpha) \ln \left\{ \frac{p(\mathbf{w}, \alpha|\mathbf{t})}{q(\mathbf{w}, \alpha)} \right\} d\mathbf{w} d\alpha \quad (10.171)$$

下界 $L(q)$ は積分できない. ここで, 局所的変分下界をロジスティックシグモイドの各因子に適応することで, 下界 $L(q)$ の下界が得られ

$$\begin{aligned} \ln p(\mathbf{t}) &\geq L(q) \geq \tilde{L}(q, \boldsymbol{\xi}) \\ &= \int\int q(\mathbf{w}, \alpha) \ln \left\{ \frac{h(\mathbf{w}, \boldsymbol{\xi}) p(\mathbf{w}|\alpha) p(\alpha)}{q(\mathbf{w}, \alpha)} \right\} d\mathbf{w} d\alpha \quad (10.172) \end{aligned}$$

となる. そして,

$$q(\mathbf{w}, \alpha) = q(\mathbf{w}) q(\alpha) \quad (10.173)$$

と分解できる.
各因子の最適解を求めると, まず因子 $q(\mathbf{w})$ の最適解は

$$\begin{aligned} \ln q(\mathbf{w}) &= E_\alpha[\ln \{h(\mathbf{w}, \boldsymbol{\xi}) p(\mathbf{w}|\alpha) p(\alpha)\}] + const \\ &= \ln h(\mathbf{w}, \boldsymbol{\xi}) + E_\alpha[p(\mathbf{w}|\alpha)] + const \\ &= -\frac{E[\alpha]}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^{N} \{(t_n - 1/2) \mathbf{w}^T \boldsymbol{\phi}_n - \lambda(\xi_n) \mathbf{w}^T \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \mathbf{w}\} + const \end{aligned}$$

これは, 因子 $q(\mathbf{w})$ がガウス分布となることが分かる.

$$q(\mathbf{w}) = N(\mathbf{w}|\boldsymbol{\mu}_N, \Sigma_N) \quad (10.174)$$

となり, ここで

$$\boldsymbol{\mu}_N = \Sigma_N \sum_{n=1}^{N} (t_n - 1/2) \boldsymbol{\phi}_n \quad (10.175), \quad \Sigma_N^{-1} = E[\alpha] I + 2 \sum_{n=1}^{N} \lambda(\xi_n) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \quad (10.176)$$

となる.
また, 因子 $q(\alpha)$ の最適解は

$$\begin{aligned} \ln q(\alpha) &= E_{\mathbf{w}}[\ln \{p(\mathbf{w}|\alpha)\}] + \ln p(\alpha) + const \\ &= \ln h(\mathbf{w}, \boldsymbol{\xi}) + E_\alpha[p(\mathbf{w}|\alpha)] + const \\ &= -\frac{M}{2} \ln \alpha - \frac{\alpha}{2} E[\mathbf{w}^T \mathbf{w}] + (a_0 - 1) \ln \alpha - b_0 \alpha + const \end{aligned}$$

これは, 因子 $q(\alpha)$ がガンマ分布となることが分かる.

$$q(\alpha) = Gam(\alpha|a_N, b_N) \quad (10.177)$$

となり, ここで

$$a_N = a_0 + \frac{M}{2} \quad (10.178)$$

$$b_N = b_0 + \frac{1}{2}E_W[W^T W] \quad (10.179)$$

となる.
また, 周辺尤度関数を直接最適化して, 前と同様に変分パラメータの更新式

$$(\xi_n^{new})^2 = \boldsymbol{\phi}_n^T(\Sigma_N + \boldsymbol{\mu}\boldsymbol{\mu}_N)\boldsymbol{\phi}_n \quad (10.181)$$

を得ることができる.

変分ロジスティック回帰

1. まずパラメータ $\boldsymbol{\xi}$ を初期化する.
2. $q(\mathbf{w}), q(\alpha)$ のパラメータを推定する.
3. $\boldsymbol{\xi}$ を推定する.
4. 2,3 を繰り返す.

3.2 コード

変分ロジスティック回帰のコード (variational_logistic_regression.py).

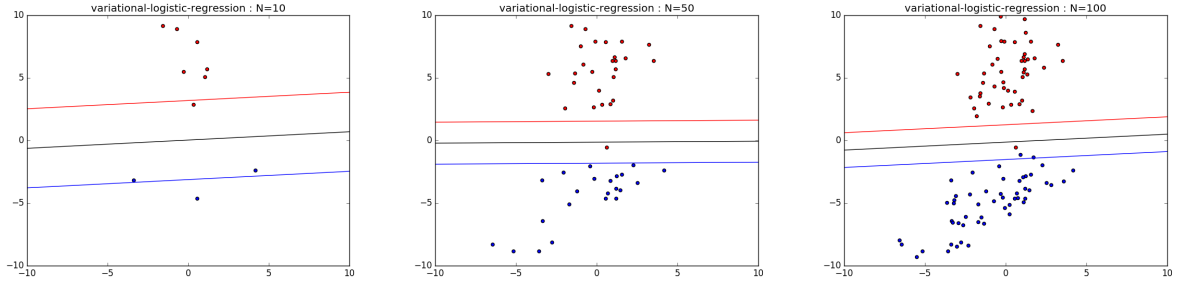
```
for N in [10,50,100]:
  x=data[:N,:2]
  #t=1 or 0
  t=data[:N,2]

  #Phi=(1,x1,x2)
  M=3
  P=np.ones((N,M))
  P[:,1:3]=x

  I=np.identity(M)
  xi=rd.rand(N)
  lam=np.zeros(N)
  mu0=np.zeros(M)
  Sig0=np.identity(M)
  a0=0
  b0=0
  muN=np.zeros(M)
  SigN=np.identity(M)
  aN=a0+M/2
  bN=b0+M/2

  diff=10**6
  while diff>=10**-6:
    diff=0
    for n in range(N):
      lam[n]=(sig(xi[n])-1/2)/(2*xi[n])

    tmp2=SigN
    tmp1=np.zeros((M,M))
    for n in range(N):
      tmp1+=lam[n]*outer(P[n,:],P[n,:])
    tmp1*=2
    SigN=inv(aN/bN*I+tmp1)
```

```
    diff+=norm(tmp2-SigN)

    tmp2=muN
    tmp1=np.zeros(M)
    for n in range(N):
      tmp1+=(t[n]-1/2)*P[n,:]
    muN=dot(SigN,tmp1)
    diff+=norm(tmp2-muN)

    tmp2=bN
    bN=b0+trace(SigN+outer(muN,muN))/2
    diff+=norm(tmp2-bN)

    tmp2=xi
    for n in range(N):
      xi[n]=sqrt(dot(P[n,:],dot(SigN+outer(muN,muN),P[n,:])))
    diff+=norm(tmp2-xi)

  def model(z):
    y=np.array([1,z[0],z[1]])
    return sig(dot(muN,y))
```

3.3 結果

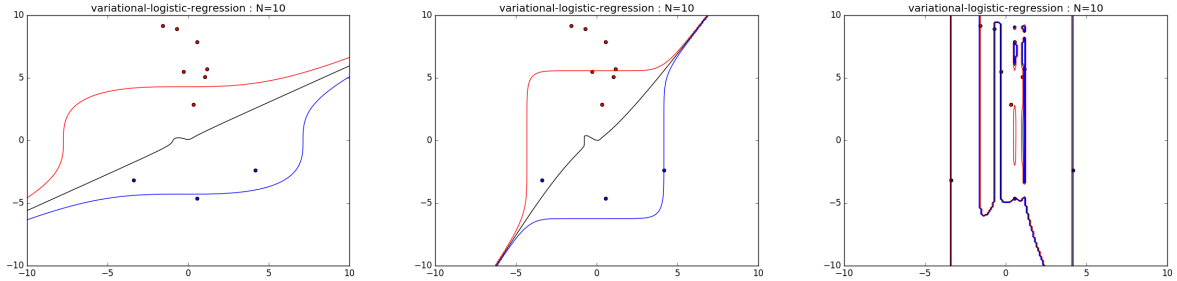
3.3.1 内積基底

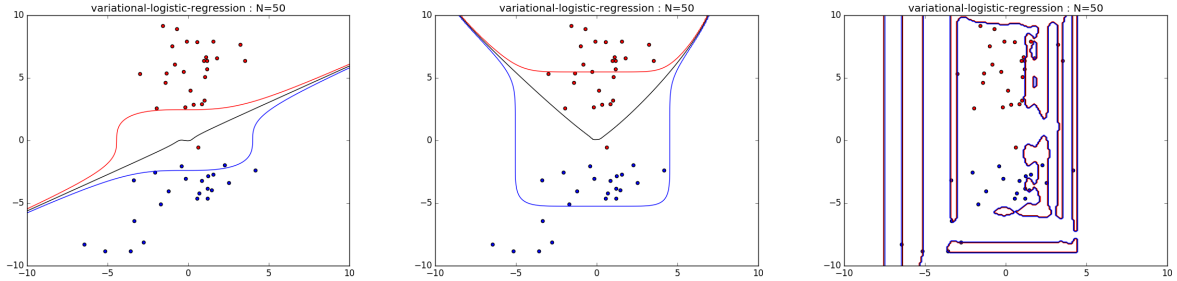
基本としてまず, 線形分離を行うために $\phi = (1, x_1, x_2)^T$ を用いた.

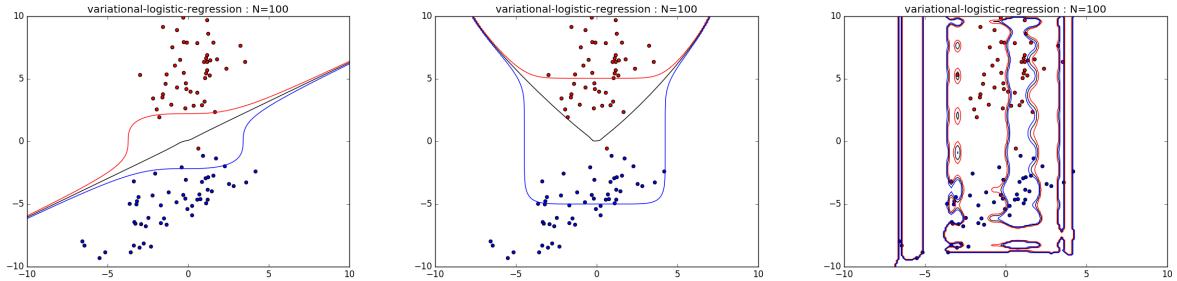


3.3.2 多項式基底

x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = x^i$ を選んだ.



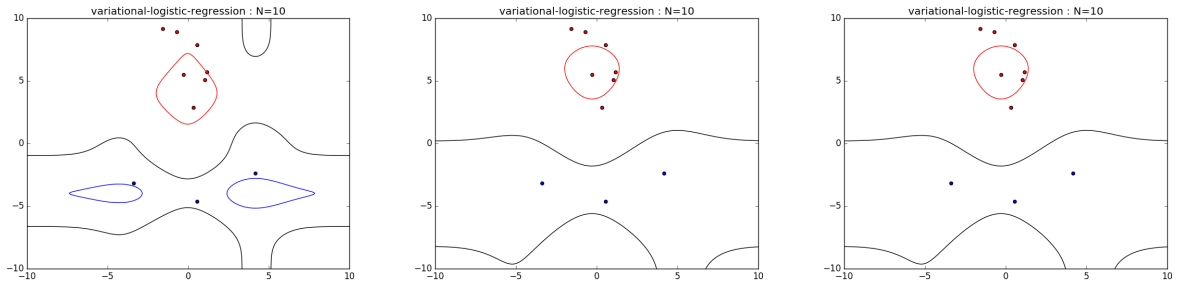


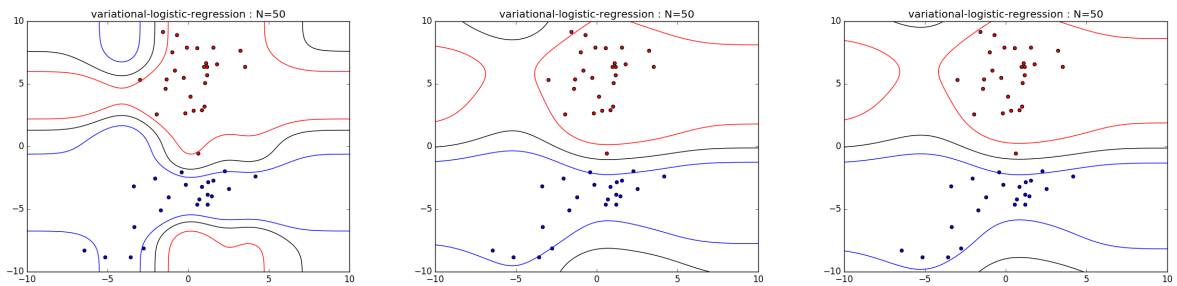


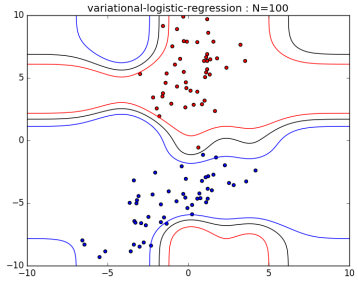
3.3.3 ガウス基底

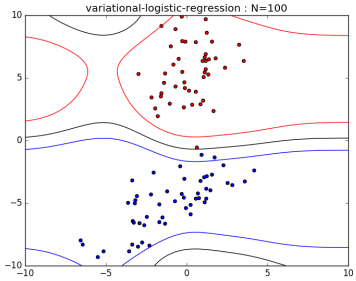
x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = exp\{-\frac{(x-\mu_i)^2}{2s^2}\}$ をもちいる．またここでは μ は区間 $[-10, 10]$ を M 個に等分する区間の中心を用い,s には 1.5 を用いた.

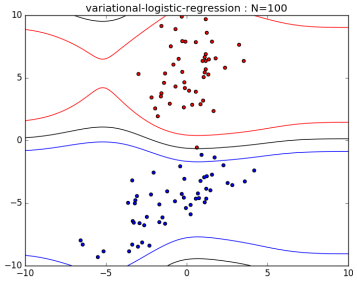
$$mu = [(m + 0.5) * (20/(M + 1)) - 10 \ for \ m \ in \ range(M)] \quad s = 1.5$$





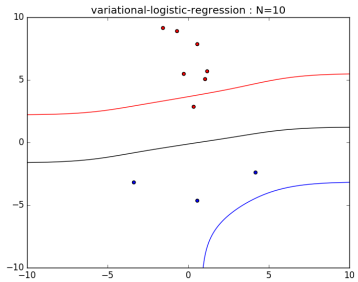


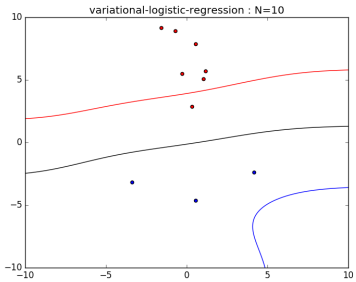


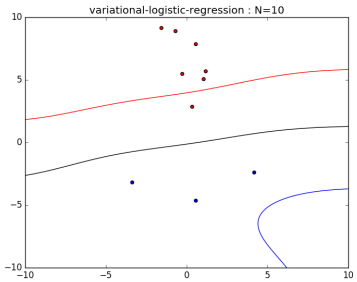


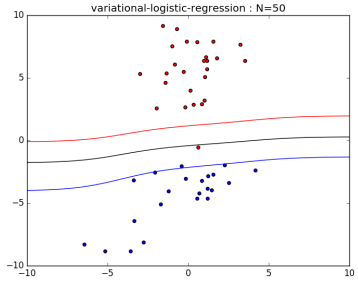
3.3.4 シグモイド基底

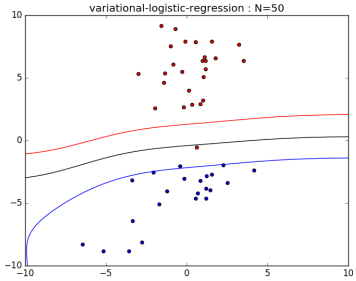
x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = \sigma(\frac{x-\mu_i}{s})$ をもちいる. ただし,$\sigma(a) = \frac{1}{1+e^{-a}}$

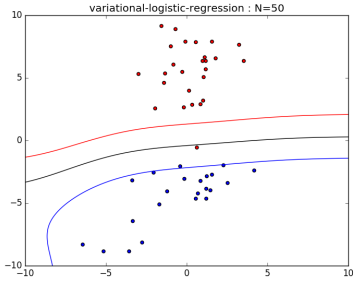


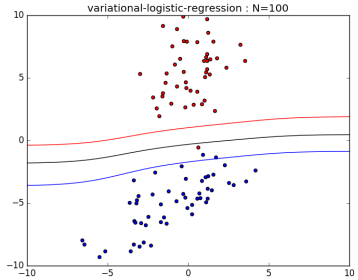


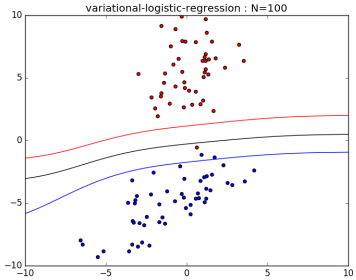


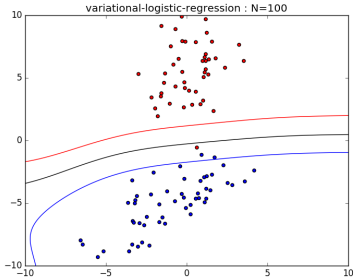






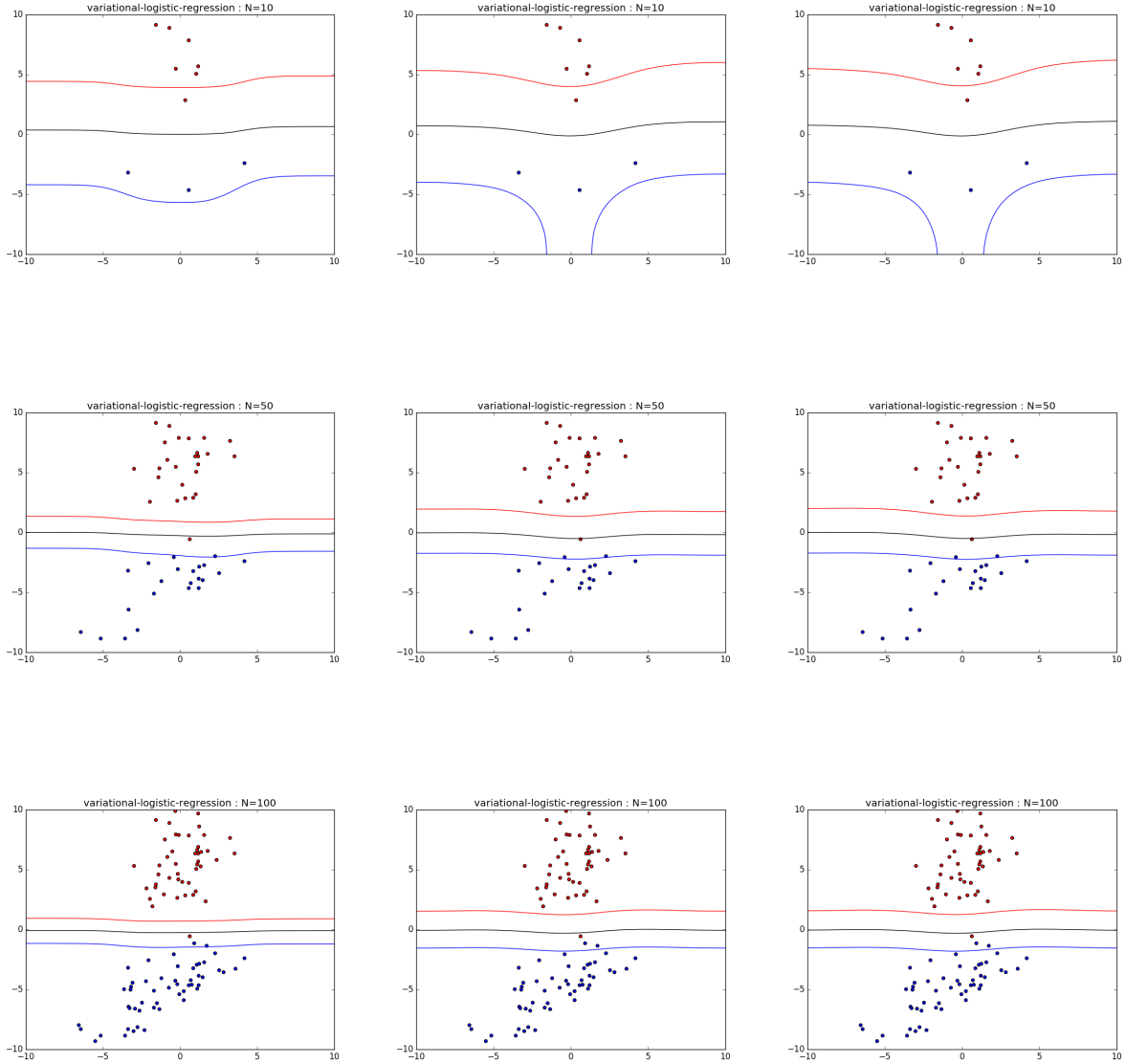






3.3.5 tanh 基底

x の各成分に対して基底 $\Phi(x)$ に $\phi_i(x) = tanh(\frac{x-\mu_i}{s})$ をもちいる.



4　まとめ

　近似とはいえかなりの精度で分類ができると分かった.
多項式基底はやはり異常なものとなった. また, ガウス基底もあまり良い結果とはならなかった.
そのほか, 内積, シグモイド, tanh 基底はうまくいった.
超パラメータを考えたときも考えなかったときも大体同じ結果となった.