Computazione per l'interazione naturale: regressione logistica Bayesiana

Corso di Interazione uomo-macchina II

Prof. Giuseppe Boccignone

Dipartimento di Informatica
Università di Milano

boccignone@di.unimi.it
http://boccignone.di.unimi.it/IUM2_2014.html

Modelli discriminativi probabilistici //Regressione Logistica

- Trasformiamo il dato di input $\mathbf{x} = [x_1, \cdots, x_D]^\mathsf{T}$ usando M funzioni di base

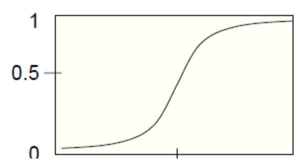
$$\phi(\mathbf{x}) = [\phi_1(\mathbf{x}, \cdots, \phi_M(\mathbf{x})]^\mathsf{T},$$

- Usiamo un modello lineare per descrivere la log-likelihood ratio

$$\log \frac{P(C=1|\mathbf{x})}{P(C=0|\mathbf{x})} = \mathbf{w}^\mathsf{T}\phi(\mathbf{x})$$

- Poichè $P(C=1|\mathbf{x}) + P(C=0|\mathbf{x}) = 1$

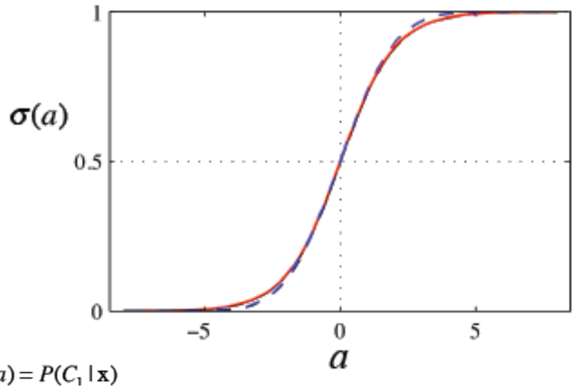
Funzione Logistica



$$\frac{P(C=1|\mathbf{x})}{1-P(C=1|\mathbf{x})} = \exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x})\right) \Rightarrow P(C=1|\mathbf{x}) = \frac{1}{1+\exp\left(-\mathbf{w}^\mathsf{T}\phi(\mathbf{x})\right)} = \frac{\exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x})\right)}{1+\exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x})\right)}$$

Modelli discriminativi probabilistici //regressione logistica: funzione logistica

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



Proprietà di simmetria
$$\sigma(-a) = 1 - \sigma(a)$$

La funzione inversa è la funzione logit
$$a = \ln\left(\frac{\sigma}{1-\sigma}\right)$$

$$\sigma(a) = P(C_1 \mid \mathbf{x})$$
$$\downarrow$$
$$a = \ln\left(\frac{\sigma}{1-\sigma}\right) \Rightarrow \ln[p(C_1|\mathbf{x})/p(C_2|\mathbf{x})]$$ Log-odds ratio

Derivata
$$\frac{d\sigma}{da} = \sigma(1-\sigma)$$

Modelli discriminativi probabilistici //Regressione Logistica

- Algoritmo di base:
- Step 1. Calcolo la funzione logit $a = \ln\left(\frac{\sigma}{1-\sigma}\right)$ con una regressione

$$\log \frac{P(C=1|\mathbf{x})}{P(C=0|\mathbf{x})} = \mathbf{w}^{\mathsf{T}}\phi(\mathbf{x})$$

- Step 2. Inverto la logit ottenendo la logistica, cioè la posteriori

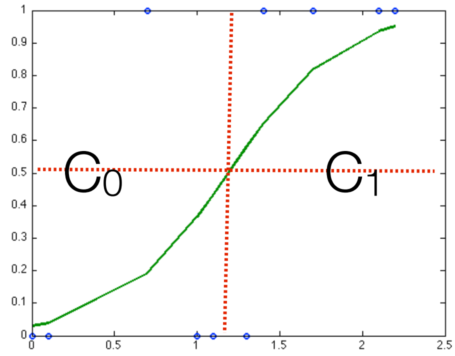
$$P(C=1|\mathbf{x}) = \frac{1}{1+\exp\left(-\mathbf{w}^{\mathsf{T}}\phi(\mathbf{x})\right)}$$

Modelli discriminativi probabilistici //regressione logistica: esempio a 2 classi

```
function EsempioLogisticRegression()
    %dati di training
    x = [0.0 0.1 0.7 1.0 1.1 1.3 1.4 1.7 2.1 2.2]';
    y = [0 0 1 0 0 0 1 1 1 1]';

    %fitting con generalized linear model dello Statistical
    %Toolbox

    w = glmfit(x,[y ones(10,1)],'binomial','link','logit')


    %predizione lineare
    %z = w(1) + x * (w(2))

    %applicazione della funzione logistica alla componente
    %lineare
    z = Logistic(w(1) + x * (w(2)))
    figure(1)
    plot(x,y,'o', x,z,'-', 'LineWidth',2)

end

function Output = Logistic(Input)

    Output = 1 ./ (1 + exp(-Input));

end
```

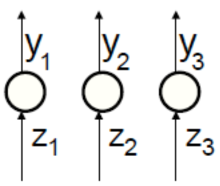
$$\log \frac{p(C_1 \mid \mathbf{x})}{p(C_0 \mid \mathbf{x})} = \mathbf{w}^T \mathbf{x} + w_0$$

$$p(C_1 \mid \mathrm{x}) = \frac{1}{1+\exp(-a)} = \sigma(a)$$



Modelli discriminativi probabilistici //regressione logistica

- Estensione a più classi: uso la decisione con funzione softmax



outputs $y_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$

$$\frac{\partial y_j}{\partial z_i} = \delta(i,j) y_j - y_j y_i$$

- La logistica è un caso particolare di softmax a due classi

$$y_1 = \frac{e^{z_1}}{e^{z_1} + e^{z_0}} = \frac{1}{1 + e^{-(z_1 - z_0)}}$$

$z' = z_1 - z_0 = w_1 x - w_2 x = wx$

Modelli discriminativi probabilistici //regressione logistica Bayesiana

- Scriviamo la probabilità congiunta

a priori

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

likelihood

- Funzione di likelihood (iid): $p(\mathbf{t}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} P(C = t_n|\mathbf{x}_n, \mathbf{w})$

- I target t_n sono binari e seguono una distribuzione di Bernoulli

$$\begin{aligned} P(C = t_n|\mathbf{x}_n, \mathbf{w}) &= P(C = 1|\mathbf{x}_n, \mathbf{w})^{t_n} \times (1 - P(C = 1|\mathbf{x}_n, \mathbf{w}))^{1-t_n} \\ &= \left[\frac{1}{1 + \exp\left(-\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_n)\right)}\right]^{t_n} \left[\frac{1}{1 + \exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_n)\right)}\right]^{1-t_n} \\ &= \frac{\exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_n)\right)^{t_n}}{1 + \exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_n)\right)} \end{aligned}$$

Modelli discriminativi probabilistici //regressione logistica Bayesiana

- Scriviamo la probabilità congiunta

a priori

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

likelihood

- Probabilità a priori: utilizziamo una Gaussiana sui coefficienti w

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

Modelli discriminativi probabilistici //regressione logistica Bayesiana

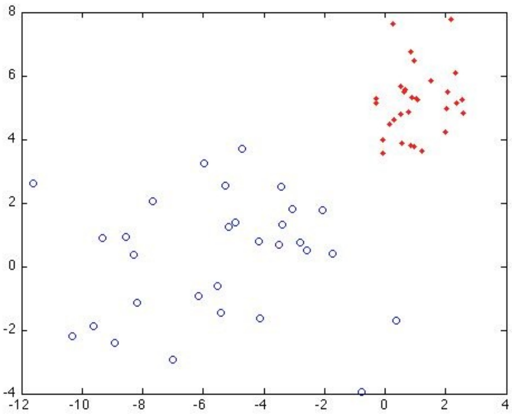
- Scriviamo la probabilità congiunta

$$p(\mathbf{t}, \mathbf{w} | \mathbf{X}, \alpha) = \underbrace{p(\mathbf{t} | \mathbf{X}, \mathbf{w})}_{\text{likelihood}} \overbrace{p(\mathbf{w} | \alpha)}^{\text{a priori}}$$

$$= \prod_{n=1}^{N} \frac{\exp\left(\mathbf{w}^{\mathsf{T}} \phi(\mathbf{x}_n)\right)^{t_n}}{1 + \exp\left(\mathbf{w}^{\mathsf{T}} \phi(\mathbf{x}_n)\right)} \mathcal{N}_{\mathbf{w}}(\mathbf{0}, \alpha^{-1}\mathbf{I})$$

Modelli discriminativi probabilistici //regressione logistica Bayesiana

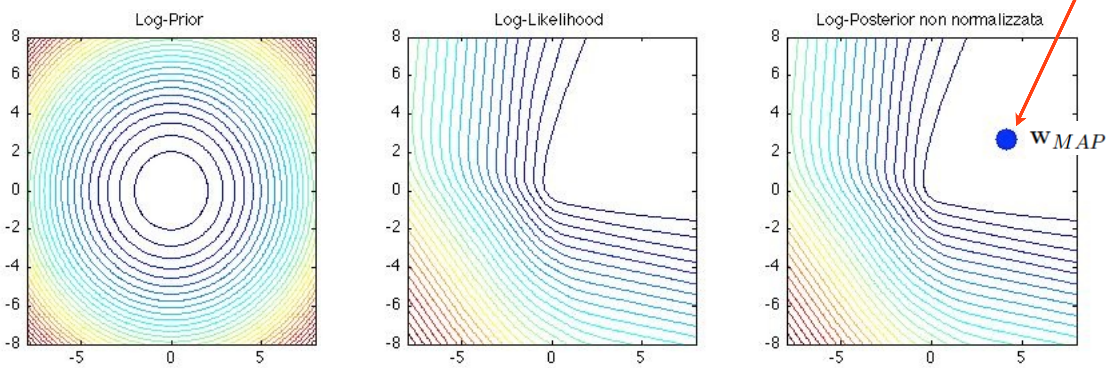
```
%Dimensione del sample e dimensione del dato
N=30;
D=2;

%Limiti e grid size per il contour plot
Range=8;
Step=0.1;

%2 classi con  2 features ciascuan distribuita  con due
Gaussiane
%con medie pari ai vettori mu1 & mu2 e covarianze isotrope
mu1=[ones(N,1) 5*ones(N,1)];
mu2=[-5*ones(N,1) 1*ones(N,1)];
class1_std = 1;
class2_std = 1.1;

%genera le features delle classi  e le target labels
X = [class1_std*randn(N,2)+mu1;2*class2_std*randn(N,2)+mu2];
%punti
t = [ones(N,1);zeros(N,1)];  %target = label 1/0

%plotta i punti labellati
figure
plot(X(find(t==1),1),X(find(t==1),2),'r.');
plot(X(find(t==0),1),X(find(t==0),2),'bo');
hold
```



Modelli discriminativi probabilistici //regressione logistica Bayesiana

```
%Definisce la contour grid dei parametri w1 & w2 per
%logistic regression
[w1,w2]=meshgrid(-Range:Step:Range,-Range:Step:Range);
[n,n]=size(w1);
W=[reshape(w1,n*n,1) reshape(w2,n*n,1)]; %matrice di tutti i pesi

%Calcola la log-prior, la likelihood,  e joint likelihood
%in ciascun punto della griglia definita sopra
alpha=100; %Varianza del prior
f=W*X';

Log_Prior = log(gauss(zeros(1,D),eye(D).*alpha,W));

Log_Like = W*X'*t - sum(log(1+exp(f)),2);

Log_Joint = Log_Like + Log_Prior;
```

```
%Identifica i parametri w1 & w2 che
massimizzano la posterior della
joint
[i,j]=max(Log_Joint);

plot(W(j,1),W(j,2),'.','MarkerSize',
40);
```



Modelli discriminativi probabilistici //regressione logistica Bayesiana

- Inferenza sui parametri (Bayes): $p(\mathbf{w}|\mathbf{t},\mathbf{X},\alpha) = p(\mathbf{t}|\mathbf{X},\mathbf{w})p(\mathbf{w}|\alpha)\dfrac{1}{p(\mathbf{t}|\mathbf{X},\alpha)}$

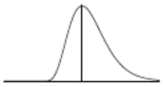
$$
\begin{aligned}
p(\mathbf{t}|\mathbf{X},\alpha) &= \int p(\mathbf{t}|\mathbf{X},\mathbf{w})p(\mathbf{w}|\alpha)d\mathbf{w} \\
&= \int \prod_{n=1}^{N} \frac{\exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_n)\right)^{t_n}}{1+\exp\left(\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_n)\right)} \mathcal{N}_{\mathbf{w}}(\mathbf{0},\alpha^{-1}\mathbf{I})d\mathbf{w}
\end{aligned}
$$

likelihood marginale: non computabile analiticamente!

- Possibilità:
  - Tecniche Monte Carlo
  - Approssimazione di Laplace
  - Tecniche variazionali di approssimazione

Modelli discriminativi probabilistici
//regressione logistica Bayesiana: Laplace

- Idea semplice: la distribuzione da approssimare ha un massimo in x0



$P^*(x)$

univariata

multivariata

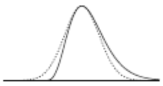
$\ln P^*(x)$

Univariata:

$$\ln P^*(x) \simeq \ln P^*(x_0) - \frac{c}{2}(x - x_0)^2 + \cdots$$

$$c = -\left.\frac{\partial^2}{\partial x^2} \ln P^*(x)\right|_{x=x_0}$$

$\ln P^*(x)$
$\& \ln Q^*(x)$



$P^*(x)$
$\& Q^*(x)$

$$Q^*(x) \equiv P^*(x_0) \exp\left[-\frac{c}{2}(x - x_0)^2\right]$$

$$Z_P \equiv \int P^*(x)\,\mathrm{d}x \quad = P^*(x_0)\sqrt{\frac{2\pi}{c}}.$$

Multivariata:

$$\ln P^*(\mathbf{x}) \simeq \ln P^*(\mathbf{x}_0) - \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathsf{T}}\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + \cdots$$

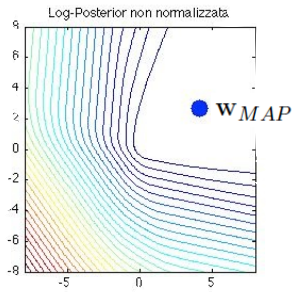
$$A_{ij} = -\left.\frac{\partial^2}{\partial x_i \partial x_j} \ln P^*(\mathbf{x})\right|_{\mathbf{x}=\mathbf{x}_0}$$

$$Z_P \simeq Z_Q = P^*(\mathbf{x}_0)\frac{1}{\sqrt{\det \frac{1}{2\pi}\mathbf{A}}} = P^*(\mathbf{x}_0)\sqrt{\frac{(2\pi)^K}{\det \mathbf{A}}}$$

Modelli discriminativi probabilistici
//regressione logistica Bayesiana: Laplace

- La posteriori $p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \alpha)$ da approssimare ha un massimo in $\mathbf{w}_{MAP}$
- Covarianza calcolata nel massimo

$$\mathbf{C} = -\left(\frac{\partial^2}{\partial \mathbf{w} \partial \mathbf{w}^{\mathsf{T}}} \log p(\mathbf{t}, \mathbf{w}|\mathbf{X}, \alpha)\right)^{-1}$$



- Approssimazione di Laplace

$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \alpha) = p(\mathbf{t}|\mathbf{X}, \mathbf{w})p(\mathbf{w}|\alpha)\frac{1}{p(\mathbf{t}|\mathbf{X}, \alpha)} \approx \mathcal{N}_{\mathbf{w}}(\mathbf{w}_{MAP}, \mathbf{C})$$

Modelli discriminativi probabilistici
//regressione logistica Bayesiana: Laplace

- Per trovare il massimo, massimizziamo la log-likelihood della congiunta

$$
\begin{aligned}
\mathcal{L} = \log p(\mathbf{t}, \mathbf{w} | \mathbf{X}, \alpha) &= \sum_{n=1}^{N} t_n \mathbf{w}^\mathsf{T} \phi(\mathbf{x}_n) \\
&- \log\left(1 + \exp\left(\mathbf{w}^\mathsf{T} \phi(\mathbf{x}_n)\right)\right) \\
&- \frac{1}{\alpha}\mathbf{w}^\mathsf{T}\mathbf{w} - \frac{D}{2}\log(2\pi\alpha^2)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{w}} &= \sum_{n=1}^{N} t_n \phi(\mathbf{x}_n) - P(C=1|\mathbf{x}_n)\phi(\mathbf{x}_n) - \frac{1}{\alpha}\mathbf{w} \\
&= \mathbf{\Phi}^\mathsf{T}\mathbf{t} - \mathbf{\Phi}^\mathsf{T}\mathbf{p} - \frac{1}{\alpha}\mathbf{w}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \mathbf{w} \partial \mathbf{w}^\mathsf{T}} &= -\sum_{n=1}^{N} \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^\mathsf{T} P(C=1|\mathbf{x}_n)\left(1 - P(C=1|\mathbf{x}_n)\right) - \frac{1}{\alpha}\mathbf{I} \\
&= -\mathbf{\Phi}^\mathsf{T}\mathbf{V}\mathbf{\Phi} - \frac{1}{\alpha}\mathbf{I}
\end{aligned}
$$

Modelli discriminativi probabilistici
//regressione logistica Bayesiana: Laplace

- Per trovare il massimo, massimizziamo la log-likelihood della congiunta

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{w}} &= \sum_{n=1}^{N} t_n \phi(\mathbf{x}_n) - P(C=1|\mathbf{x}_n)\phi(\mathbf{x}_n) - \frac{1}{\alpha}\mathbf{w} \\
&= \mathbf{\Phi}^\mathsf{T}\mathbf{t} - \mathbf{\Phi}^\mathsf{T}\mathbf{p} - \frac{1}{\alpha}\mathbf{w}
\end{aligned}
$$

$$
\mathbf{p} = [P(C=1|\mathbf{x}_1), \cdots, P(C=1|\mathbf{x}_N)]^\mathsf{T}
$$

$$
\mathbf{\Phi} = \begin{bmatrix} \phi_1(\mathbf{x}_1) & \cdots & \phi_M(\mathbf{x}_1) \\ \vdots & \phi_m(\mathbf{x}_n) & \vdots \\ \phi_1(\mathbf{x}_N) & \cdots & \phi_M(\mathbf{x}_N) \end{bmatrix}
$$

Modelli discriminativi probabilistici //regressione logistica Bayesiana: Laplace

- Per trovare il massimo, massimizziamo la log-likelihood della congiunta

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial \mathbf{w} \partial \mathbf{w}^\mathsf{T}} &= -\sum_{n=1}^{N} \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^\mathsf{T} P(C=1|\mathbf{x}_n)\left(1 - P(C=1|\mathbf{x}_n)\right) - \frac{1}{\alpha}\mathbf{I} \\ &= -\mathbf{\Phi}^\mathsf{T}\mathbf{V}\mathbf{\Phi} - \frac{1}{\alpha}\mathbf{I}\end{aligned}$$

$$\begin{bmatrix} v_{11} & 0 & \cdots & 0 \\ 0 & v_{22} & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & v_{NN} \end{bmatrix} \qquad v_{nn} = P(C=1|\mathbf{x}_n)(1 - P(C=1|\mathbf{x}_n))$$

- Dall'Hessiana scrivo la matrice di covarianza della distribuzione approssimata

$$\mathbf{C} = \left(\mathbf{\Phi}^\mathsf{T}\mathbf{V}\mathbf{\Phi} + \frac{1}{\alpha}\mathbf{I}\right)^{-1}$$

Modelli discriminativi probabilistici //regressione logistica Bayesiana: Laplace

- Quasi fatta...

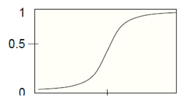
$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \alpha) = p(\mathbf{t}|\mathbf{X}, \mathbf{w})p(\mathbf{w}|\alpha)\frac{1}{p(\mathbf{t}|\mathbf{X}, \alpha)} \approx \mathcal{N}_\mathbf{w}(\mathbf{w}_{MAP}, \mathbf{C})$$

$$\mathbf{C} = \left(\mathbf{\Phi}^\mathsf{T}\mathbf{V}\mathbf{\Phi} + \frac{1}{\alpha}\mathbf{I}\right)^{-1}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{w}} &= \sum_{n=1}^{N} t_n\phi(\mathbf{x}_n) - P(C=1|\mathbf{x}_n)\phi(\mathbf{x}_n) - \frac{1}{\alpha}\mathbf{w} \\ &= \mathbf{\Phi}^\mathsf{T}\mathbf{t} - \mathbf{\Phi}^\mathsf{T}\mathbf{p} - \frac{1}{\alpha}\mathbf{w} \quad = 0\end{aligned}$$

i termini di p sono
non lineari
non computabile
analiticamente!

$$\mathbf{p} = [P(C=1|\mathbf{x}_1), \cdots, P(C=1|\mathbf{x}_N)]^\mathsf{T}$$

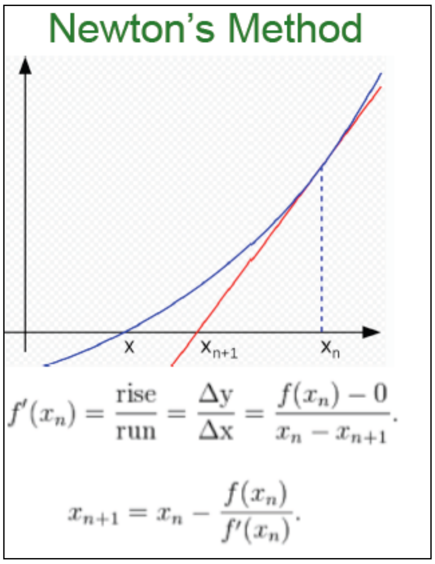


Modelli discriminativi probabilistici //regressione logistica Bayesiana: Laplace

- Trovo lo zero con un metodo approssimato iterativamente: IRLS (iterative Reweighted Least Square) una variante del metodo di Newton- Raphson

$$\mathbf{w} \leftarrow \mathbf{w} - \left(\frac{\partial^2 \mathcal{L}}{\partial \mathbf{w} \partial \mathbf{w}^\mathsf{T}} \right)^{-1} \frac{\partial \mathcal{L}}{\partial \mathbf{w}}$$

$$\begin{aligned}
\mathbf{w} \quad \leftarrow & \quad \mathbf{w} + \mathbf{C} \left(\mathbf{\Phi}^\mathsf{T} \mathbf{t} - \mathbf{\Phi}^\mathsf{T} \mathbf{p} - \frac{1}{\alpha} \mathbf{w} \right) \\
= & \quad \mathbf{C} \left(\mathbf{C}^{-1} \mathbf{w} + \mathbf{\Phi}^\mathsf{T} \mathbf{t} - \mathbf{\Phi}^\mathsf{T} \mathbf{p} - \frac{1}{\alpha} \mathbf{w} \right) \\
= & \quad \left(\mathbf{\Phi}^\mathsf{T} \mathbf{V} \mathbf{\Phi} + \frac{1}{\alpha} \mathbf{I} \right)^{-1} \mathbf{\Phi}^\mathsf{T} \left(\mathbf{V} \mathbf{\Phi} \mathbf{w} + \mathbf{t} - \mathbf{p} \right)
\end{aligned}$$



Modelli discriminativi probabilistici //regressione logistica Bayesiana: Laplace

- Soluzione finale

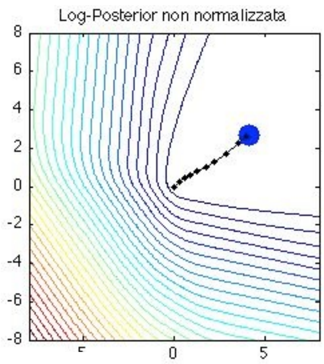
$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \alpha) = p(\mathbf{t}|\mathbf{X}, \mathbf{w}) p(\mathbf{w}|\alpha) \frac{1}{p(\mathbf{t}|\mathbf{X}, \alpha)} \approx \mathcal{N}_\mathbf{w}(\mathbf{w}_{MAP}, \mathbf{C})$$

$$\mathbf{C} = \left(\mathbf{\Phi}^\mathsf{T} \mathbf{V} \mathbf{\Phi} + \frac{1}{\alpha} \mathbf{I} \right)^{-1}$$

$$\mathbf{w} \quad \leftarrow \quad \left(\mathbf{\Phi}^\mathsf{T} \mathbf{V} \mathbf{\Phi} + \frac{1}{\alpha} \mathbf{I} \right)^{-1} \mathbf{\Phi}^\mathsf{T} \left(\mathbf{V} \mathbf{\Phi} \mathbf{w} + \mathbf{t} - \mathbf{p} \right)$$

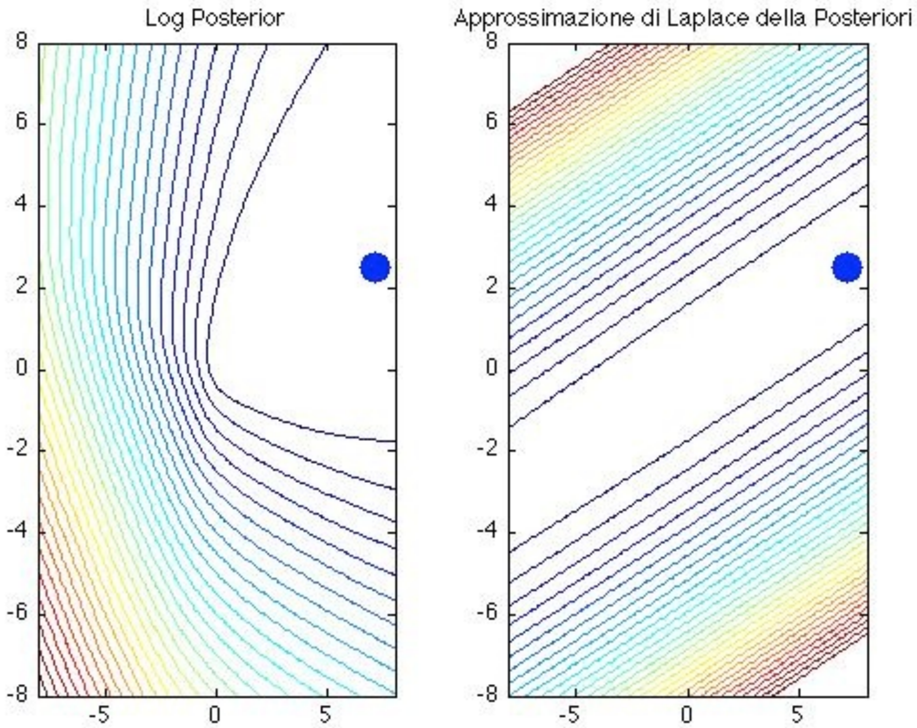
Esempio Matlab

```
%Newton routine per trovare i valori MAP  di w1 & w2
%Fisso un numero di step a 10 e stime iniziali a w1=0, w2=0
N_Steps = 10;
w = [0;0];

for m=1:N_Steps
    %memorizza i parametri e plotta l'evoluzione delle stime
    ww(m,:) = w;
    plot(ww(:,1),ww(:,2),'k.-');
    drawnow
    pause(0.1)

    %Newton Step
    P = 1./(1 + exp(-X*w));
    V = diag(P.*(1-P));
    H = inv(X'*V*X + eye(D)./alpha);

    w = H*X'*(V*X*w + t - P);

    %Calcola le nuove likelihood
    f=X*w;
    lpr = log(gauss(zeros(1,D),eye(D).*alpha,w'));
    llk = f'*t - sum(log(1+exp(f)));
    ljt = llk + lpr;
    fprintf('Log-Likelihood = %f, Joint-Likelihood = %f\n',llk,ljt)
end
```



$$\mathbf{w} \leftarrow \left(\mathbf{\Phi}^{\mathsf{T}}\mathbf{V}\mathbf{\Phi} + \frac{1}{\alpha}\mathbf{I}\right)^{-1} \mathbf{\Phi}^{\mathsf{T}}\left(\mathbf{V}\mathbf{\Phi}\mathbf{w} + \mathbf{t} - \mathbf{p}\right)$$

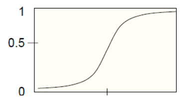
Modelli discriminativi probabilistici //regressione logistica Bayesiana: Laplace



Modelli discriminativi probabilistici //regressione logistica Bayesiana: Classificazione

- In termini Bayesiani: classificazione = predizione

$$P(C=1|\mathbf{x}_{new},\alpha,\mathbf{X},\mathbf{t}) = \int P(C=1|\mathbf{x}_{new},\mathbf{w})p(\mathbf{w}|\mathbf{X},\mathbf{t},\alpha)d\mathbf{w}$$



non computabile analiticamente!

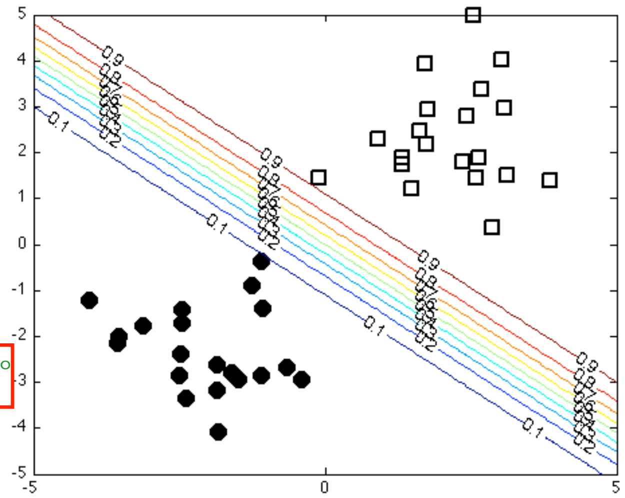
- Metodo 1: assumo che sia fortemente "piccata" nel valore MAP

$$\begin{aligned} P(C=1|\mathbf{x}_{new},\alpha,\mathbf{X},\mathbf{t}) &\approx P(C=1|\mathbf{x}_{new},\mathbf{w}_{MAP},\alpha,\mathbf{X},\mathbf{t}) \\ &= \frac{1}{1+\exp(-\mathbf{w}_{MAP}^{\mathsf{T}}\phi(\mathbf{x}_{new}))} \end{aligned}$$

$$P(C=1|\mathbf{x}_{new},\alpha,\mathbf{X},\mathbf{t}) > 0.5 \quad \Rightarrow \quad C=1$$

Modelli discriminativi probabilistici //regressione logistica Bayesiana: Classificazione

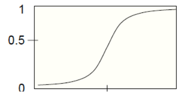
```
%% Inizializza i parametri
w = repmat(0,2,1); % posizioniamo in 0

tol = 1e-6; % tolleranza per fermare l'iterazione
Nits = 100;
w_all = zeros(Nits,2); % Evoluzione dei parametri verso il massimo
ss = 10; % varianza della Prior sui parametri  w
change = inf;
it = 0;
while change>tol & it<=100
    prob_t = 1./(1+exp(-X*w));
    % Gradiente
    grad = -(1/ss)*w' + sum(X.*(repmat(t,1,length(w))-repmat(prob_t,1,length(w))),1);
    % Hessiano
    H = -X'*diag(prob_t.*(1-prob_t))*X;
    H = H - (1/ss)*eye(length(w));
    % Update di w
    w = w - inv(H)*grad';
    it = it + 1;
    w_all(it,:) = w';
    if it>1
        change = sum((w_all(it,:) - w_all(it-1,:)).^2);
    end
end
w_all(it+1:end,:) = [];


%% Approssimazione MAP

[Xv,Yv] = meshgrid(-5:0.1:5,-5:0.1:5); %definisce il supporto
Probs = 1./(1+exp(-(w(1).*Xv + w(2).*Yv)));
```



Modelli discriminativi probabilistici
//regressione logistica Bayesiana: Classificazione

- In termini Bayesiani: classificazione = predizione

$$P(C=1|\mathbf{x}_{new}, \alpha, \mathbf{X}, \mathbf{t}) = \int P(C=1|\mathbf{x}_{new}, \mathbf{w}) p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha) d\mathbf{w}$$



non computabile analiticamente!

- Metodo 2: risolvo l'integrale con un Metodo Montecarlo:
  - campiono un numero Ns di parametri:

$$\mathbf{w}_s \sim \mathcal{N}(\mathbf{w}_{MAP}, \mathbf{C})$$

  - approssimo l'integrale come media della somma

$$\begin{aligned} P(C=1|\mathbf{x}_{new}, \alpha, \mathbf{X}, \mathbf{t}) &\approx \frac{1}{N} \sum_{n_s=1}^{N_s} P(C=1|\mathbf{x}_{new}, \mathbf{w}_s) \\ &= \frac{1}{N} \sum_{n_s=1}^{N_s} \frac{1}{1+\exp(-\mathbf{w}_s^{\mathsf{T}} \phi(\mathbf{x}_{new}))} \end{aligned}$$