Lecture 22

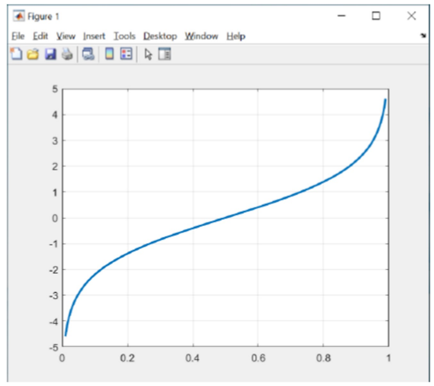
Logistic Regression

logit function

The sigmoid function:

$$\sigma(a) = p(C_1|\mathbf{x}) = \frac{1}{1+e^{-a}}$$

where $a(\mathbf{x}) = \ln\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})}$



The **logit** function (also called the **log odd** function) represents the log of the ratio of probabilities:

$$a(\mathbf{x}) = \ln\left(\frac{\sigma}{1-\sigma}\right) = \ln\frac{p(C_1|\mathbf{x})}{1-p(C_1|\mathbf{x})} = \ln\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})}$$

```
>> s = 0.01: 0.001: 0.99;
>> a = log(s./(1-s));
>> plot(s,a); grid
```

The **logistic** (sigmoid) function and the **logit** function are inverse of each other.

Softmax Function

Multiclass ($K > 2$) generalization of the logistic sigmoid to a *normalized exponential*:

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k)}{\sum_j p(\mathbf{x}|\mathcal{C}_j)p(\mathcal{C}_j)} = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

where $a_k = \ln p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k)$

$$e^{a_k} = e^{\ln p(x|C_k)p(C_k)} = p(x|C_k)p(C_k)$$

In contrast to the two-class case:

$$p(\mathcal{C}_1|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)+p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)} = \frac{1}{1+\exp(-a)} = \sigma(a)$$

$$a = \ln\frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)}$$

The *softmax* function represents a smoothed version of the "max" function because, if $a_k \gg a_j$ for all $j \neq k$, then $p(C_k|\mathbf{x}) \approx 1$, and $p(C_j|\mathbf{x}) \approx \mathbf{0}$.

Generalized Linear Models and Link Function

- So far we have considered classification models that work directly with the original input vector $\mathbf{x}$.
- We can also make a fixed nonlinear transformation of the inputs using a vector of **basis functions** $\phi(\mathbf{x})$.
- The resulting decision boundaries will be linear in the feature space Φ, and these correspond to nonlinear decision boundaries in the original $\mathbf{x}$ space.
- We begin our treatment of generalized linear models by considering the problem of two-class classification.
- Extension of logistic sigmoid function representation of the posterior probability from $\sigma(a) = p(C_1|\mathbf{x}) = \frac{1}{1+e^{-a}}$, where $\mathrm{a}(\mathbf{x}) = \ln\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})}$ is the *logic* function, to **Logistic Regression** as follows: $p(C_1|\phi) = y(\phi) = \sigma(\mathbf{w}^{\mathrm{T}}\phi)$, and $p(C_2|\phi) = 1 - p(C_1|\phi)$.
- $\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}) = \sigma^{-1}[y(\phi)] = \mathrm{a}[p(C_1|\phi(\mathbf{x}))]$. The inverse of the sigmoid – the *logit* function is called the **link function**, which converts the probability of the response variables to a generalized linear combination of explanatory variables (input vector $\mathbf{x}$).
- We have seen an example of logistic regression previously, when we fitted Gaussian class conditional densities.

linear
$\phi(x) = -1.2745x_1 - 1.7747x_2 + 19.0582$
(x_1, x_2)

$p(C_1|\phi) = \frac{1}{1+e^{-}} = y(\phi) = \sigma(?)$

- In a two-class classification problem, the posterior probability of class C_1 can be written as a logistic sigmoid acting on a linear function of the feature vector ϕ so that $p(C_1|\phi) = y(\phi) = \sigma(\mathbf{w}^{\mathrm{T}}\phi)$

mnrval ( ) function and Decision Boundary

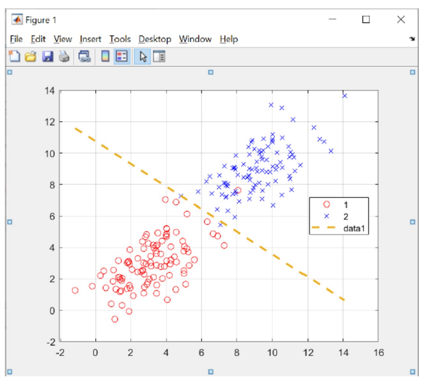
```
>> B
B =
   19.0582
   -1.2745
   -1.7747

>> x = mean(data)
x =
   6.1741    5.9980

>> prob = mnrval(B, x)
prob =
   0.6329    0.3671

>> B(2)*x(1) + B(3)*x(2) + B(1)
ans =
   0.5447

>> log(prob(1)/prob(2))
ans =
   0.5447
```




```
% Boundary: B(2)*x1 + B(3)*x2 + B(1) = 0
% log(P(C1|X)/P(C2|X)) >= 0, if P(C1|X)>=P(C2|X)

x1 = min(data(:,1)): 0.01: max(data(:,1));
x2 =  -(B(2)*x1 + B(1))/B(3);
plot(x1,x2)
```

Parametric Form for $p(C_k|\mathbf{x})$

- Assume that the class-conditional densities are Gaussian.
- We consider first two classes, and assume that all classes share the same covariance matrix.
- Thus the density for class C_k is given by

$$p(\mathbf{x}|\mathcal{C}_k) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right\}$$

$$\sigma(a) = p(C_1|\mathbf{x}) = \sigma\left(\ln\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})}\right) = \sigma\left(\ln\frac{\overbrace{p(C_1|\mathbf{x})\boldsymbol{p}(\mathbf{x})}^{p(C_1,x)}}{\underbrace{p(C_2|\mathbf{x})\boldsymbol{p}(\mathbf{x})}_{p(C_2,x)}}\right) = \sigma\left(\ln\frac{\overbrace{p(\mathbf{x}|C_1)\boldsymbol{p}(C_1)}^{p(C_1,x)}}{\underbrace{p(\mathbf{x}|C_2)\boldsymbol{p}(C_2)}_{p(C_2,x)}}\right)$$

$$p(\mathcal{C}_1|\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0)$$ where

$$\begin{aligned}
\mathbf{w} &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
w_0 &= -\frac{1}{2}\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(\mathcal{C}_1)}{p(\mathcal{C}_2)}
\end{aligned}$$

'

logistic_regression_demo.m

```
% Decision boundary based on coefficients in B
% B(2)*x1 + B(3)*x2 + B(1) = 0, where B(1) is the intercept
% log(P(C1|X)/P(C2|X)) >= 0, if P(C1|X)>=P(C2|X)
x1 = min(data(:,1)): 0.01: max(data(:,1));
x2 =  -(B(2)*x1 + B(1))/B(3);
hold on;
plot(x1,x2)
```

```
>> B

B =

   19.0582
   -1.2745
   -1.7747
```

```
>> % Slopes and intercepts based on theoretical results
inv(Sigma)*(mu1-mu2)'
-0.5*mu1*inv(Sigma)*mu1'+0.5*mu2*inv(Sigma)*mu2'

ans =
  -1.2374
  -1.7664


ans =

  18.2342
```

$$\boxed{p(\mathcal{C}_1|\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0)}$$

where the weight and bias are based on the means and covariance matrix estimated by the MLE method – too many parameters to estimate!

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(\mathcal{C}_1)}{p(\mathcal{C}_2)}$$

```
>> % Determine the class probability
x = mean(data)
prob = mnrval(B, x)
log(prob(1)/prob(2))
B(2)*x(1)+B(3)*x(2)+B(1)

x =

   6.1741    5.9980


prob =

   0.6329    0.3671


ans =

   0.5447
```

$\leftarrow \ln \frac{P(C_1|x)}{P(C_2|x)}$