Linear Models for Classification (part 2)

T-61.6020: Machine Learning, Basic Principles

Antti Sorjamaa
Time Series Prediction Group
Laboratory of Computer and Information Science
Helsinki University of Technology

Outline

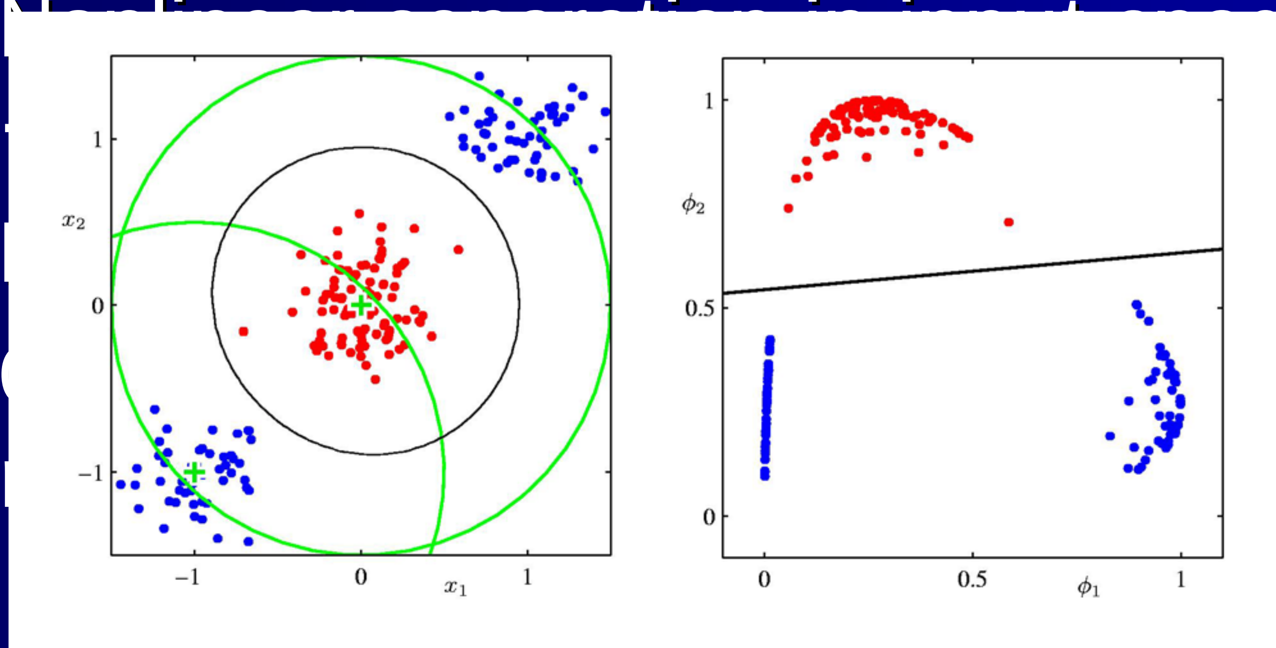
- Probabilistic Discriminative Models
  - Logistic Regression
  - Iterative Reweighted LS
  - Multiclass Logistic Regression
  - Probit Regression
- Laplace Approximation
- Bayesian Logistic Regression

Probabilistic Discriminative Models

- Direct approach
  - Maximize the conditional probability directly without going through prior and likelihood
- Fewer parameters to be determined
- Better predictive performance
- Enables the usage of basis functions

Nonlinear Basis Functions

- Transform input space to feature space through basis function



Logistic Regression

- Classification to two classes using logistic sigmoid function
- Linearly increasing number of parameters, not quadratic like Gaussians
- For linearly separable data, overfitting might occur
  - Needs regularization
  - Or prior and MAP solution

Iterative Reweighted LS

- No closed-form solution for logistic regression
  - Nonlinearity of logistic Sigmoid
  - Iterative procedure needed
- Error function has unique minimum
- Efficient algorithm: Newton-Raphson
  - Iterative, using local approximation of log likelihood

IRLS (2)

- Ordinary linear regression

$$\mathbf{w}^{new} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}$$

  – Exact solution, same as OLS

- IRLS

$$\mathbf{w}^{new} = \mathbf{w}^{old} - (\mathbf{\Phi}^{\mathrm{T}}\mathbf{R}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}(\mathbf{y}-\mathbf{t})$$

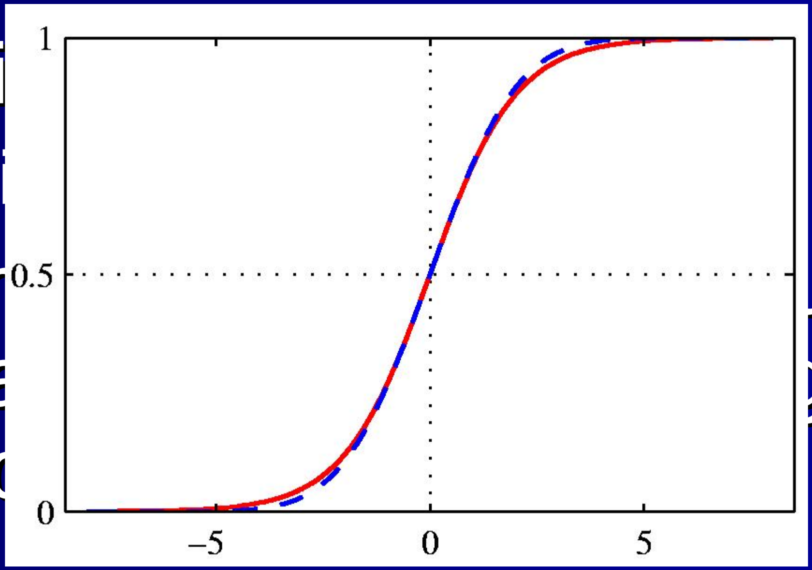
Multiclass Logistic Regression

- Two-class → Sigmoid Function
- Multiclass → Softmax Function
- IRLS algorithm used again
- Error function nearly the same than with 2-class case

$$\sigma(a) = \frac{1}{1+e^{-a}} \qquad y_k(\phi) = \frac{e^{a_k}}{\sum_j e^{a_j}}$$

Probit Regression

- Very close to the logistic regression
- Probit function used approximating Sigmoid
- Difference i
  - More sensi
- In case of r ust be taken into a g the probabilistic



Probit (2)

- Probit function

$$\Phi(a) = \int_{-\infty}^{a} N(\theta \mid 0,1)\,\mathrm{d}\theta$$

$$\Phi(a) = \frac{1}{2}\left\{1 + \frac{1}{\sqrt{2}}\mathrm{erf}(a)\right\}$$

$$\mathrm{erf}(a) = \frac{2}{\sqrt{\pi}}\int_{0}^{a} \exp\left(\frac{-\theta^2}{2}\right)\mathrm{d}\theta$$

Canonical Link Functions

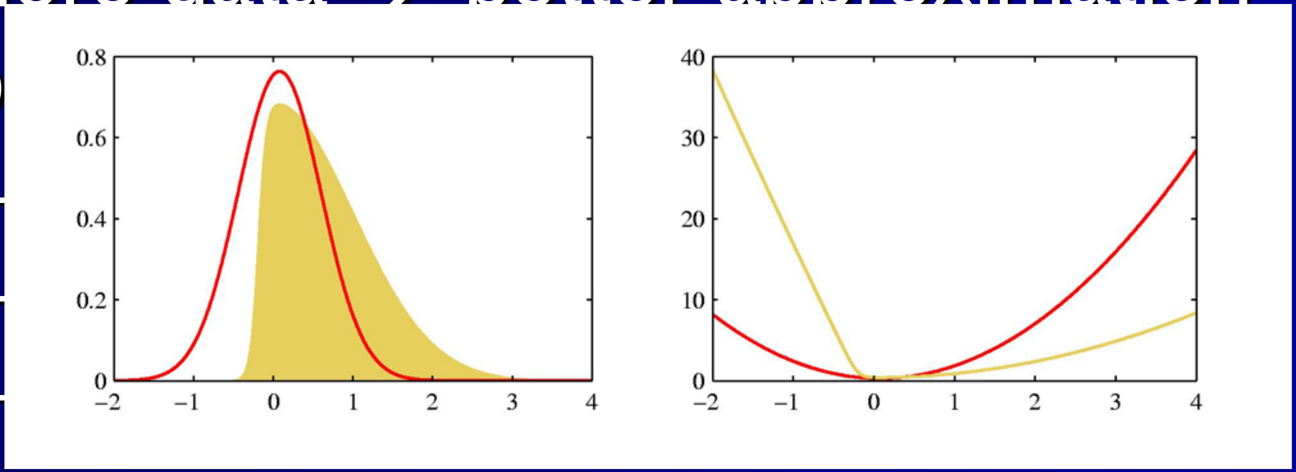
- Conditional distribution of target variable from exponential family
- Corresponding choice for activation function

  → Error formula always similar

$$\nabla \ln E(w) = \frac{1}{s} \sum_{n=1}^{N} (y_n - t_n) \phi_n$$

Laplace Approximation

- Approximation of the integration
  - Posterior no longer Gaussian
- Built around mode of the distribution
- More data → better approximation
- D



Bayesian Information Criterion (BIC)

- Same kind of model complexity measure than for example AIC
- Assumptions
  – Prior distributions broad
  – Hessian matrix full rank
  – Assumptions not always valid!
- Approximation rough
  – More accuracy in Section 5.7

BIC (2)

- BIC

$$\ln p(D) \cong \ln p(D \mid \theta_{MAP}) - \frac{1}{2} M \ln N$$

  - N data points
  - M parameters
- Penalizes model complexity harder than AIC

Bayesian Logistic Regression

- Integration of a product of Sigmoids is intractable → no exact solution
- Laplace approximation used
- Predictive distribution
  - Using Probit function to replace the Sigmoids

Bayesian Logistic Regression (2)

- Posterior distribution (approximation)

$$q(\mathbf{w}) = N\left(\mathbf{w} \mid \mathbf{w}_{MAP}, \mathbf{S}_N\right)$$

$$\mathbf{S}_N = \mathbf{S}_0^{-1} + \sum_{n=1}^{N} y_n(1-y_n)\phi_n\phi_n^{\mathrm{T}}$$

- Predictive distribution (approximation)

$$p(C_n \mid \phi, \mathbf{t}) = \sigma\left(\kappa\left(\sigma_a^2\right)\mu_a\right)$$

$$\kappa\left(\sigma^2\right) = \frac{1}{\sqrt{1+\frac{\pi\sigma^2}{8}}}$$

- Questions?

Antti.Sorjamaa@hut.fi

Pictures of the presentation, courtesy of C.M. Bishop
http://research.microsoft.com/~cmbishop/PRML/webfigs.htm