

# Statistical Methods of Data Analysis

Ulrich Husemann

Humboldt-Universität zu Berlin  
Winter Semester 2010/2011

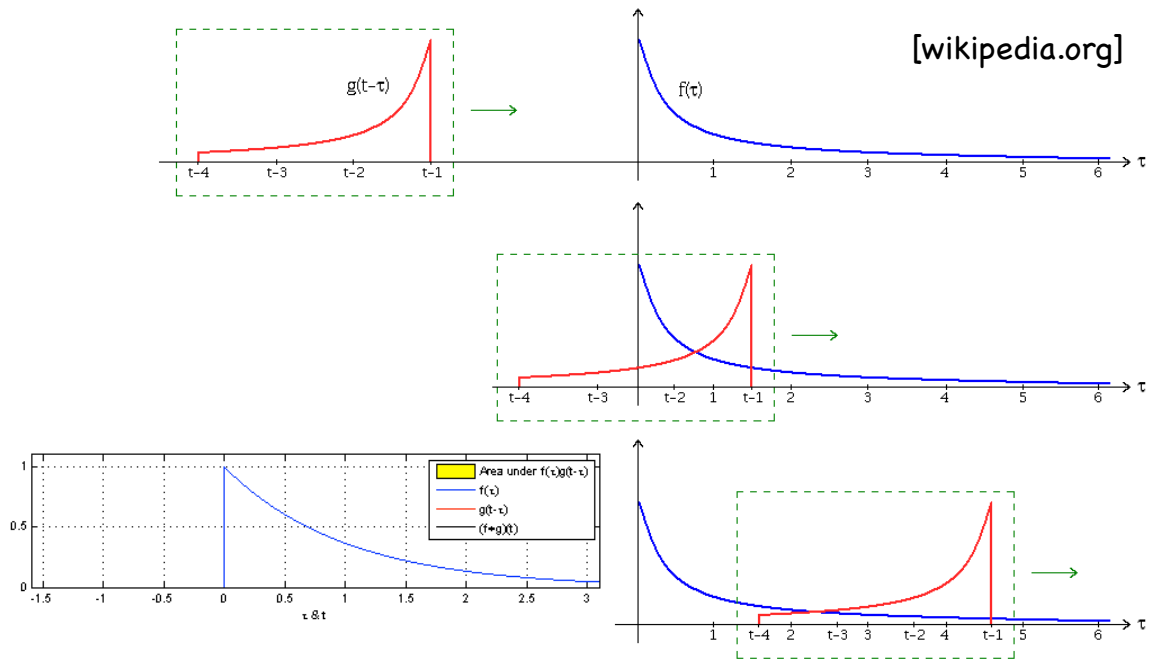
## Characteristic Functions



- Examples of characteristic functions:

- Uniform:  $f(x; a, b) = \frac{1}{b-a}$  for  $x \in [a, b]$   $\phi(k) = \frac{\exp[ikb] - \exp[ika]}{ik(b-a)}$
- Poisson:  $P(n; \nu) = \frac{\nu^n}{n!} \exp[-\nu]$   $\phi(k) = \exp[\nu(\exp[ik] - 1)]$
- Exponential:  $f(x; \lambda) = \lambda \exp[-\lambda x]$   $\phi(k) = \frac{\lambda}{\lambda - ik}$
- Gaussian:  $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$   $\phi(k) = \exp\left[i\mu k - \frac{\sigma^2 k^2}{2}\right]$
- Chi-Square:  $f(x; n) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left[-\frac{x}{2}\right]$   $\phi(k) = (1 - 2ik)^{-\frac{n}{2}}$

# Convolution

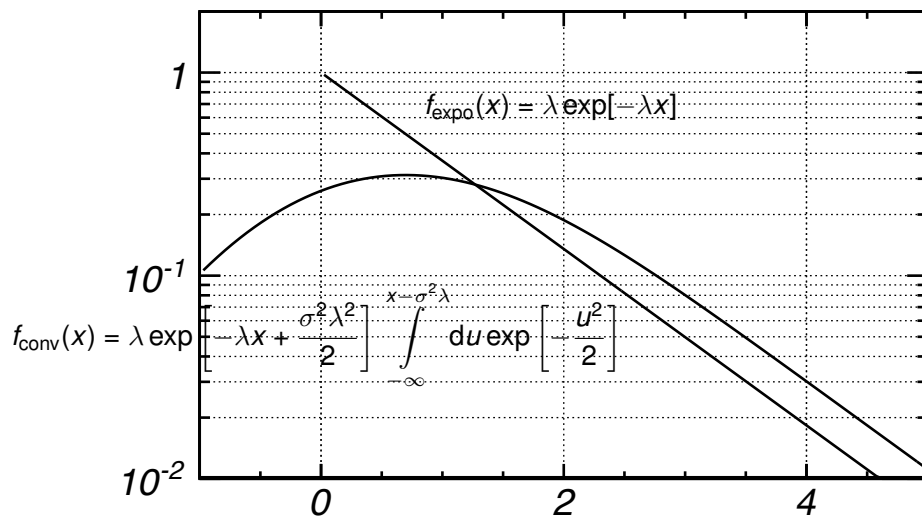


3

# Convolution



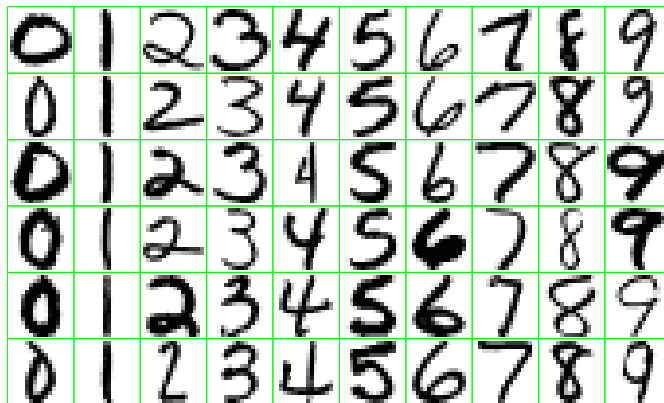
## Exponential-Gaussian Convolution



Larger contributions to convolution integral from the left  
 → convolution shifted to the right

4

# Recognition of ZIP Codes

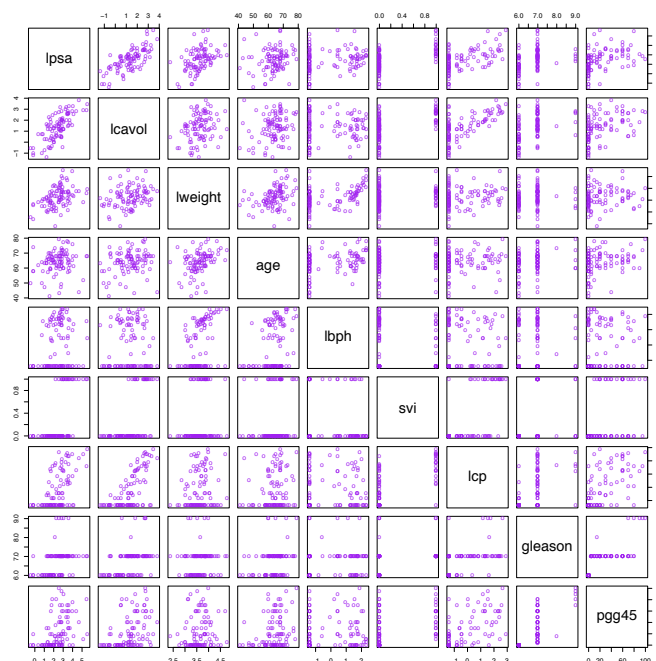


- Goal: Recognize hand-written ZIP code digits from 16x16 pixel matrix
- "Classification problem": Results can be assigned to 10 different groups (0, ..., 9)

**FIGURE 1.2.** Examples of handwritten digits from U.S. postal envelopes.

Elements of Statistical Learning (2nd Ed.), © Hastie, Tibshirani & Friedman 2009

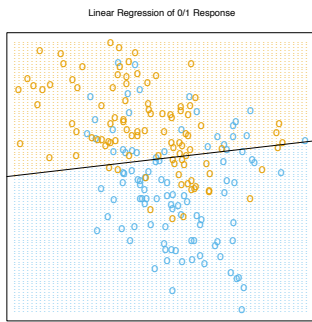
# Prostate Cancer Data



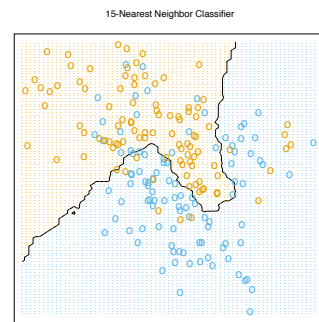
- Correlations between prostate specific antigen (PSA) and other clinical measurements
- Want: prediction of PSA from other measurements
- "Regression problem": quantitative result

Elements of Statistical Learning (2nd Ed.), © Hastie, Tibshirani & Friedman 2009

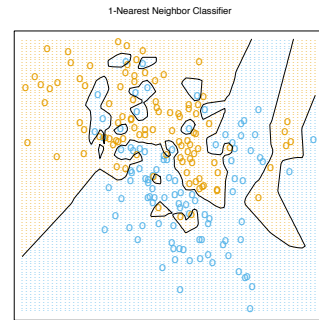
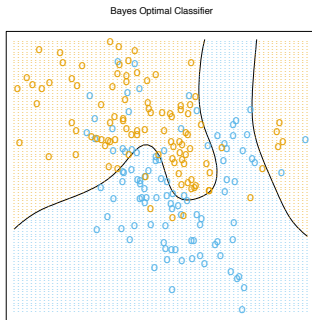
# Example Classifiers



- Classification problem: separate ORANGE from BLUE



- Various classifiers: linear, non-linear, nearest neighbors, ...



Elements of Statistical Learning (2nd Ed.), © Hastie, Tibshirani & Friedman 2009

Statistical Methods of Data Analysis (P23.1.2b), HU Berlin, WS 2010/2011, Lecture #14

7

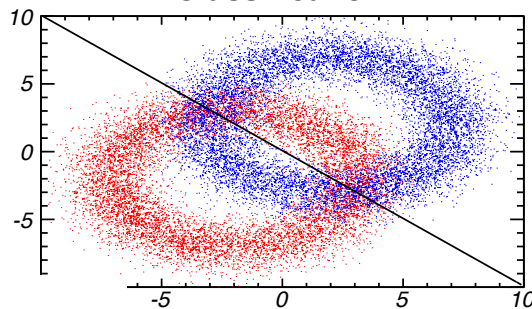
# Fisher Discriminant



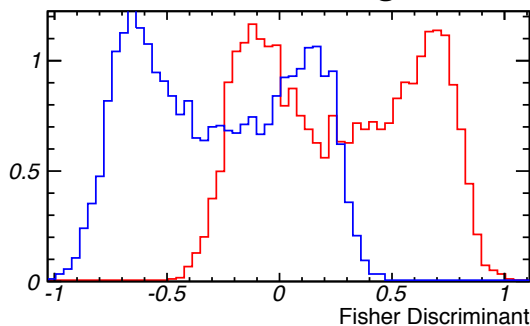
- TMVA Fisher discriminant:

- Fisher x:  $-0.067$
- Fisher y:  $-0.067$
- Fisher offset:  $+0.005$

Classification



Fisher\_tr\_S\_original



Statistical Methods of Data Analysis (P23.1.2b), HU Berlin, WS 2010/2011, Lecture #14

8