

Pattern Recognition with Hidden Markov Modells

Dynamic Programming at its Best



Univ. Doz. Dr. Stefan Wegenkittl

Fachhochschule Salzburg, Studiengang
Informationstechnik & System-Management

Stochastic Pattern Recognition

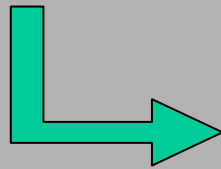
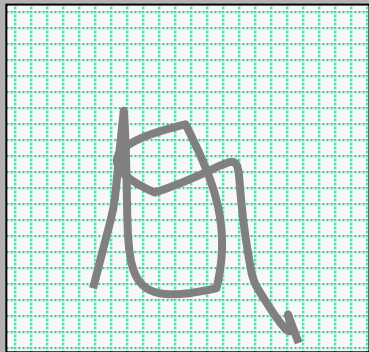
- ▶ Sir Isaiah Berlin: „To understand is to perceive patterns“
- ▶ Patterns: recognizeable, obvious entities
- ▶ Input: raw data (image, signal, time series)
- ▶ Features: unique characteristics of patterns
- ▶ Model: assumptions on relationship btw. patterns & features
- ▶ Important: find set of features which
 - ▶ reduces amount/dimension of input data
 - ▶ still contains all information necessary to distinguish patterns



Example

Inputvektor

256x256

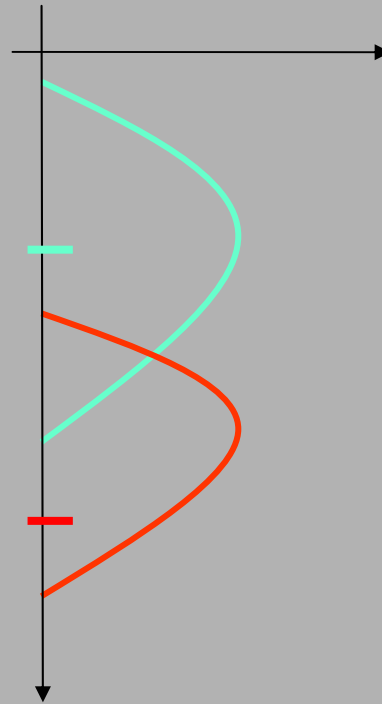


Featurevektor

%



h/l

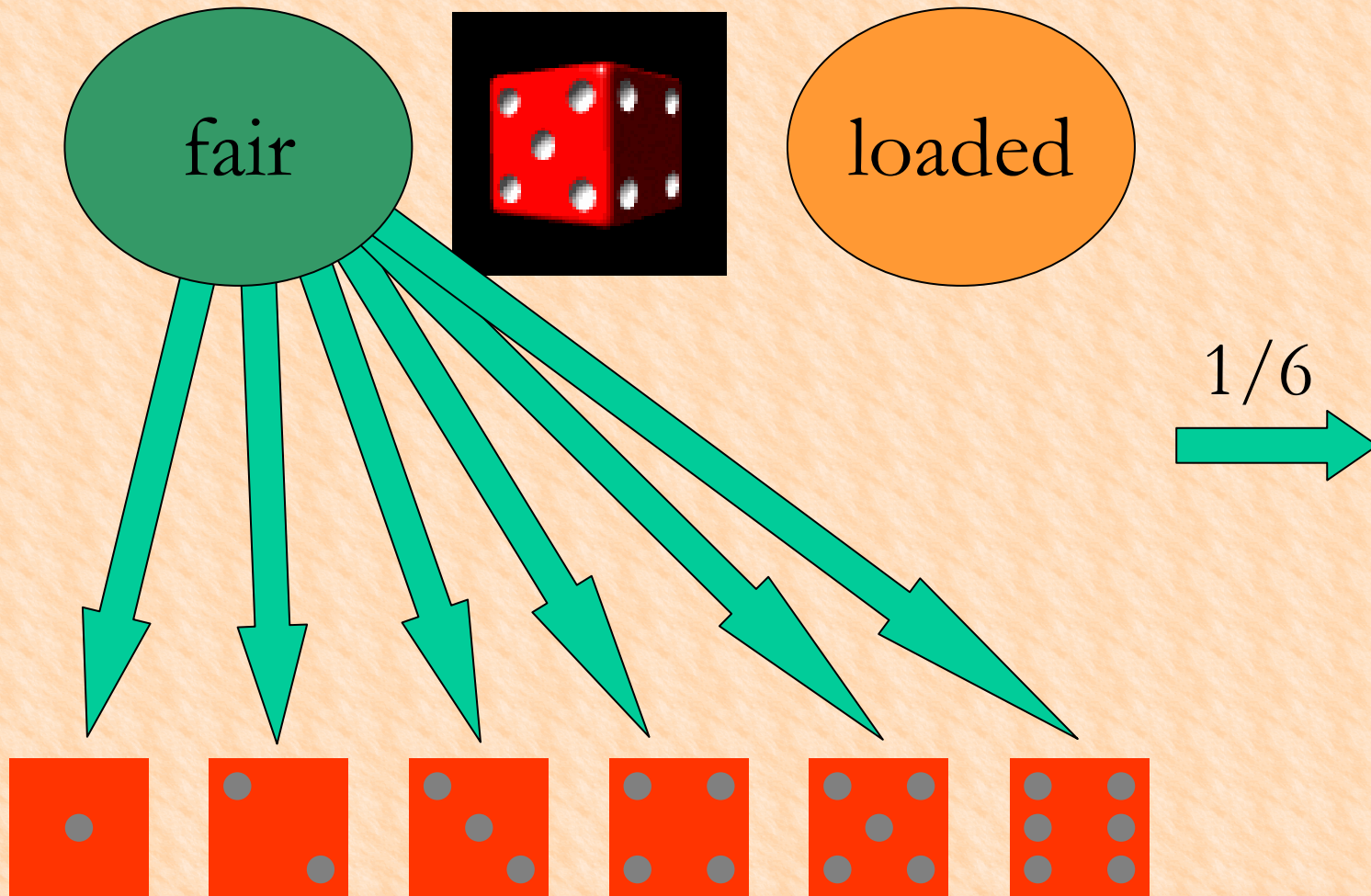


Stochastic Pattern Recognition

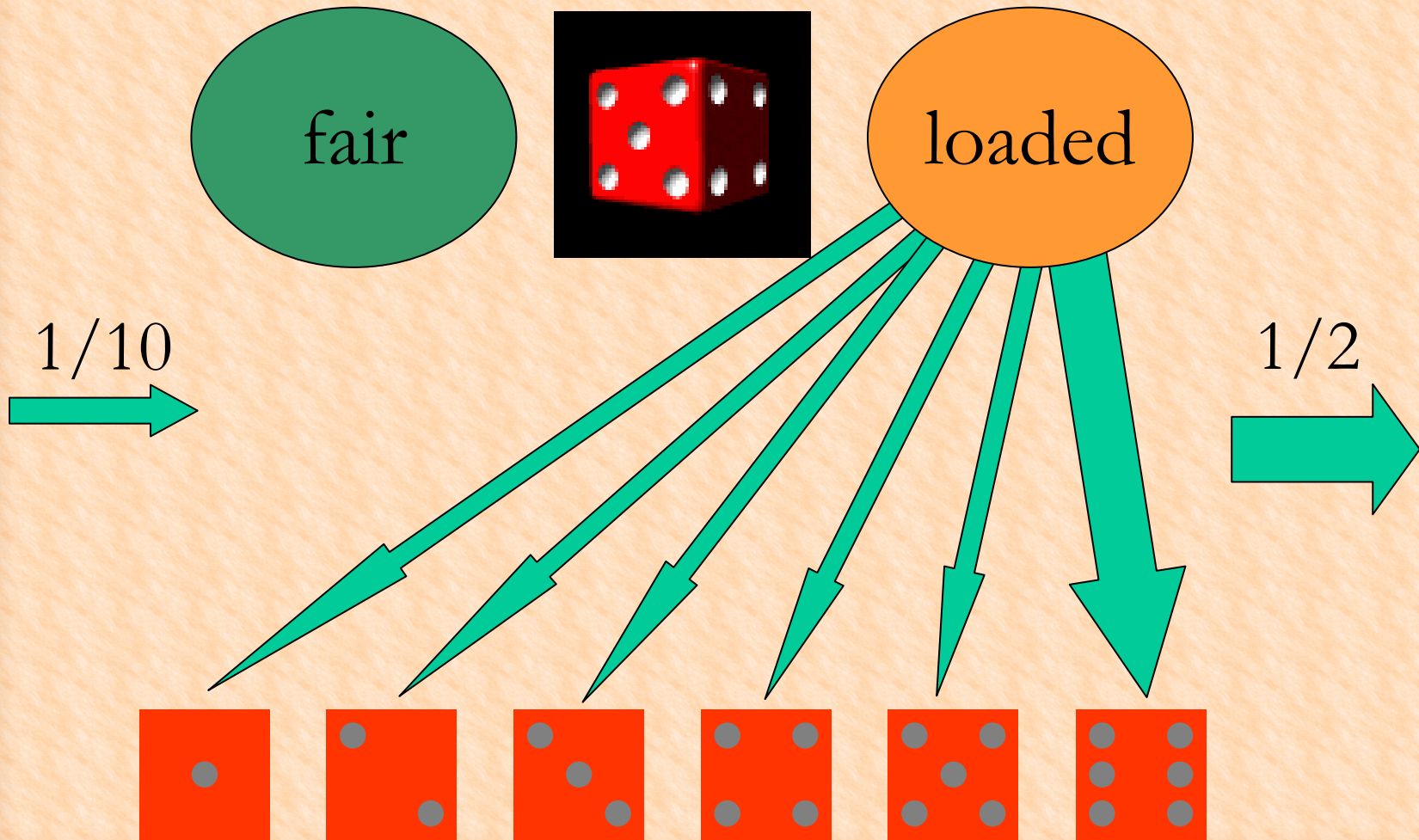
- ▶ HMMs: stochastic models for temporal/serial data
- ▶ i.e.:
 - ▶ model gives $P[\text{pattern} \mid \text{features}]$
 - ▶ features: series of natural numbers
- ▶ Applications
 - ▶ speech recognition
 - ▶ bioinformatics (gene hunting)
 - ▶ fault detection in machinery
 - ▶ DOS watchdogs
 - ▶ medical signal processing



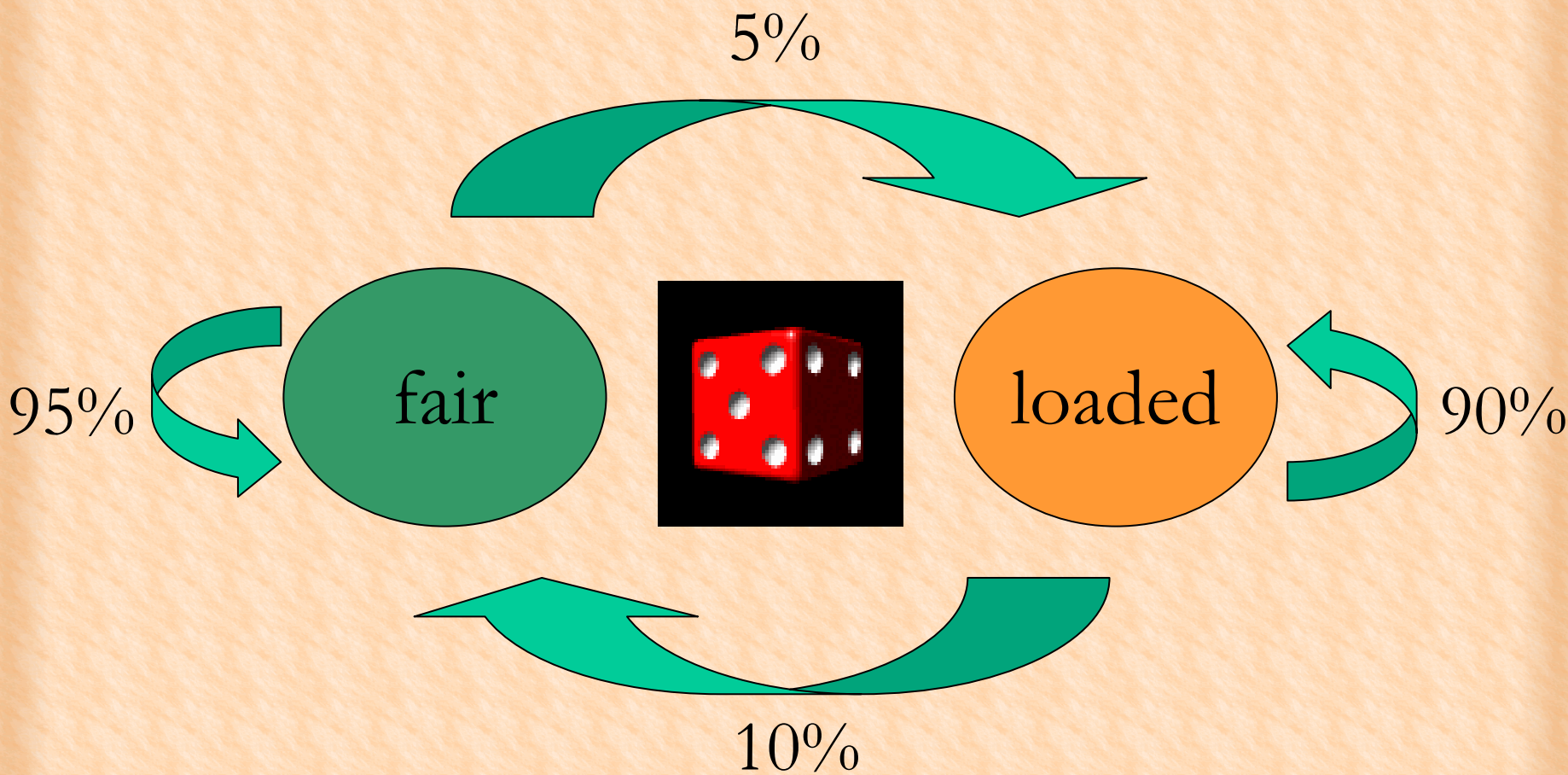
Example: The occasionally dishonest casino



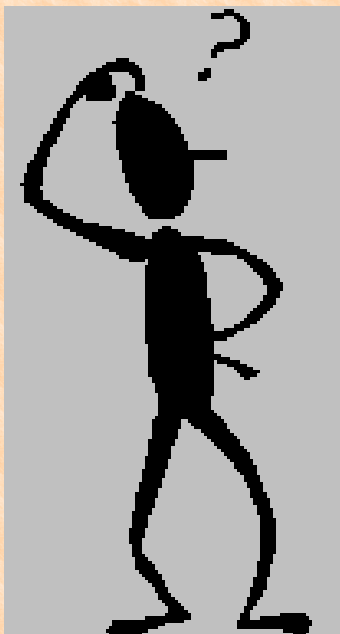
Example: The occasionally dishonest casino



Example: The occasionally disshonest casino



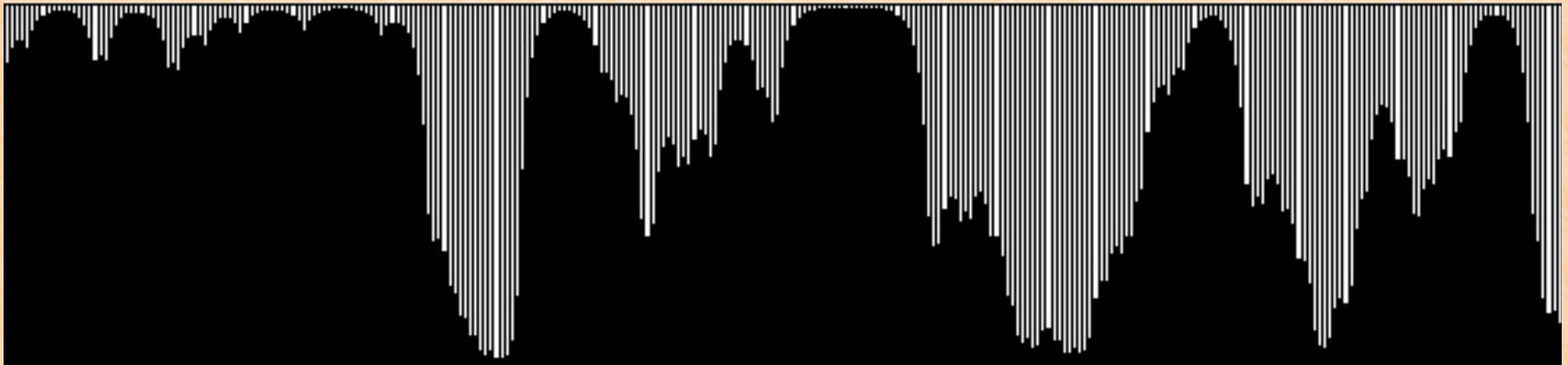
Pattern Recognition Problem



- ▶ given a list of dice rolls (outcomes)
- ▶ decide:
 - ▶ when was fair dice used
 - ▶ when loaded one
- ▶ “decoding” problem

Stochastic Model Generates Probabilities

- ▶ solid bars: probability of fair dice
- ▶ white rectangles: probability of loaded dice
- ▶ decision rule: pattern with highest probability is „recognized“



A little bit of history...



Andrei Andreyevich Markov

Born: 14 June 1856

Ryazan, Russia

Died: 20 July 1922

Petrograd, Russia



A little bit of history...

Markov is particularly remembered for his study of **Markov chains**, sequences of **random variables** in which the future variable is **determined by the present variable** but is **independent** of the way in which the present state arose from its **predecessors**.

Example: upper part of our model: coin choosing



A little bit of history...

Markov is particularly remembered

chains, sequences of random variables

variable is determined by the present

of the way in which the present

predecessors.

Example: upper part of our model: coin choosing

easy: independent coin flips (multiply probs.)

hard: arbitrary dependence between events (common distribution)

compromise: future depends on present but not on past



A little bit of history...



Thomas Bayes

Born: 1702

London, England

Died: 17 April 1761,
Kent, England



A little bit of history...

Bayes set out his **theory of probability** in *Essay towards solving a problem in the doctrine of chances* in **1764**.

Bayes's conclusions were **accepted by Laplace** in a **1781** memoir, rediscovered by Condorcet, and remained unchallenged until

Boole questioned them in the *Laws of Thought*.

Since then **Bayes' techniques** have been **subject to controversy**.



HMMs are Markov Chains

- ▶ States cannot be observed directly
- ▶ Instead: at each instance a dice (selection depends on state) is rolled, only the outcome thereof can be observed
- ▶ State is „hidden“: hidden markov model
- ▶ Mathematically: subclass of markov models, nothing new...



HMMs are Markov Chains

- ▶ 1966-77: Baum et. al.: basic theory
- ▶ 1975ff: speech recognition using HMMs
- ▶ 1985ff: good tutorials (L. R. Rabiner, ...) application to other areas, esp. Bioinformatics
- ▶ why HMMs?
 - ▶ rich math. structure => wide of range applications
 - ▶ „work well in practice“ (Rabiner, 1988)



The 3 classical problems

▶ Evaluation Problem:

- ▶ Given observation, compute probability under given HMM
- ▶ Used in Bioinformatics: model selection
- ▶ **forward algorithm**

▶ Decoding Problem:

- ▶ Given observation, compute most probable path in given HMM
- ▶ Used in Speech Recognition: find series of phonemes
- ▶ **Viterby algorithm, forward-backward algorithm**

▶ Training Problem:

- ▶ Given a lot of observations, find suitable HMM
- ▶ EM Algorithm, ...



Forward Algorithm

- ▶ Problem: complexity of computation: $O(\#S^n)$

$$P[B] = \sum_{\substack{\text{all paths } A \\ \text{of length } n}} P[B \cap A]$$

- ▶ Forward Algorithm solves this in $O(n \#S^2)$
- ▶ exploits Markov property



Viterby & Forward-Backward Algorithm

- ▶ 2 conditions for optimality:

$$A^* = \operatorname{argmax} P[B \cap A] \quad \text{Viterby}$$

A path of length n

$$A_i^* = \operatorname{argmax}_{a \in S} P[a_i = a | B] \quad \text{F.B.}$$



Remarks

▶ Viterby

- ▶ also linear in observation length n
- ▶ backtracking
- ▶ global optimal solution

▶ Forward Backward

- ▶ also linear in observation length n
- ▶ no backtracking needed
- ▶ local optimal solutions, resulting path has smaller prob. than Viterby



Bioinformatics revisited

- ▶ PFAM collection of proteins online:

- ▶ <http://pfam.wustl.edu/cgi-bin/getdesc?name=Globin>
- ▶ <http://www.rcsb.org/pdb/>
- ▶ <http://pfam.wustl.edu/hmmsearch.shtml>

- ▶ try it out:

- ▶ >1BZZ:A HEMOGLOBIN ALPHA CHAIN

```
MLSPADKTNVKAAWGKVG AHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK  
KVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPA  
VHASLDKFLASVSTVLTSKYR
```



Links & References

- ▶ Tutorial: Rabiner (pdf on www)
- ▶ further links...: www.stefan.wegenkittl.com

Danke für die Einladung!

