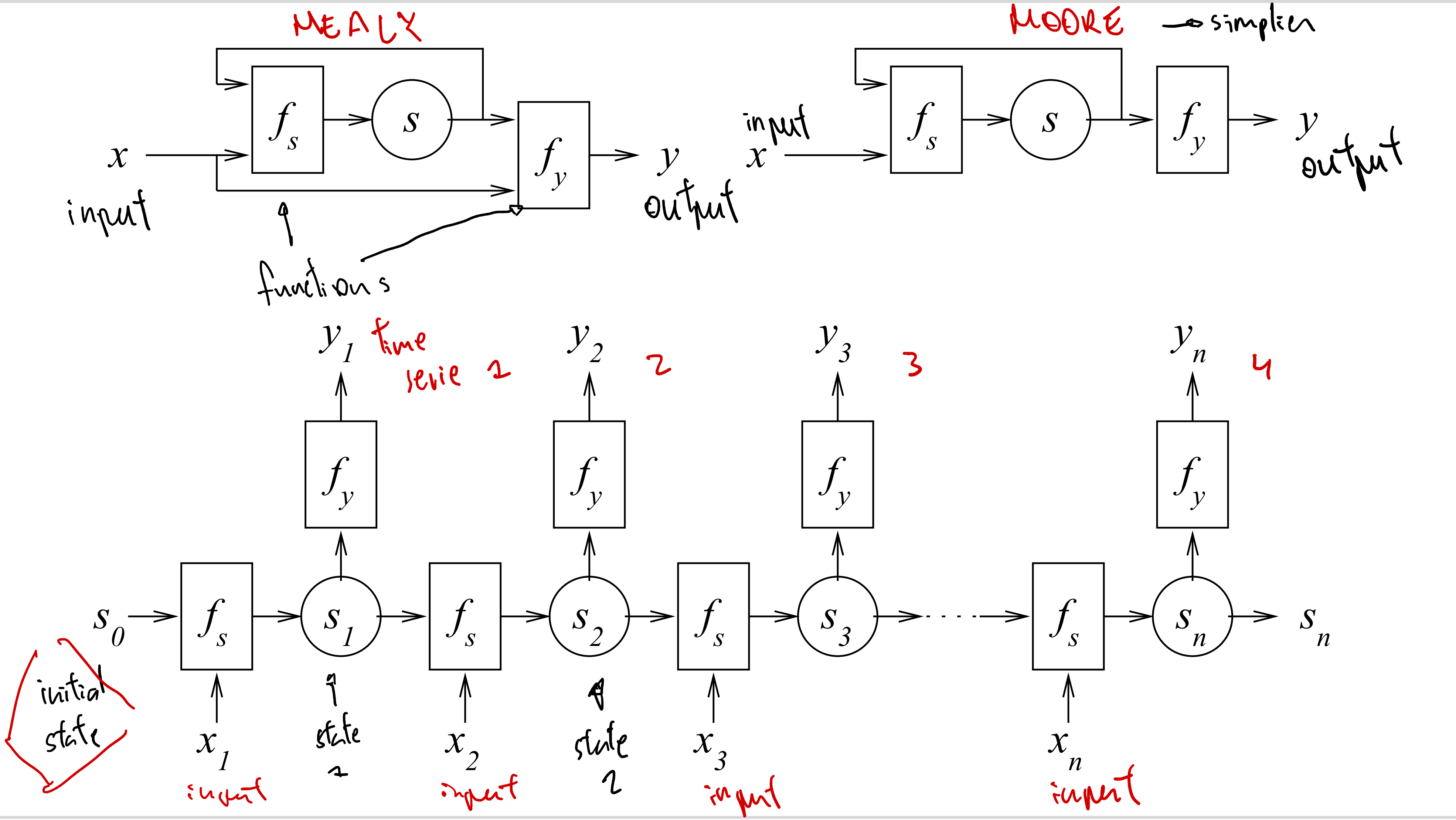
Chapter 7: Time Series Forecasting

1. Mealy and Moore Machines
2. Recurrent Models
3. Autoregressive Models

Final state machines

Mealy and Moore Machines



Prof. Dr. Thomas A. Runkler

Mealy and Moore Machines

- Mealy machine

$$
\begin{aligned}
s_k &= f_s(s_{k-1}, x_k) \\
y_k &= f_y(s_k, x_k)
\end{aligned}
$$

- Moore machine

$$
\begin{aligned}
s_k &= f_s(s_{k-1}, x_k) \\
y_k &= f_y(s_k)
\end{aligned}
$$

Recurrent Models

- recurrent model without explicit state — initial state as 0

$$y_k = f_k(y_1, \ldots, y_{k-1}, x_1, \ldots, x_{k-1}), \quad k = 2, \ldots, n$$

- constant time horizon

$$y_k = f(\overbrace{y_{k-m}, \ldots, y_{k-1}}^{\text{just first } m \text{ outputs}}, \underbrace{x_{k-m}, \ldots, x_{k-1}}_{\text{first } m \text{ inputs}}), \quad k = m+1, \ldots, n$$

- forecast model by regression with

$$\begin{aligned}
y_4 &= f(y_1, y_2, y_3, x_1, x_2, x_3) \quad \text{regression sample}\\
y_5 &= f(y_2, y_3, y_4, x_2, x_3, x_4)\\
y_6 &= f(y_3, y_4, y_5, x_3, x_4, x_5)\\
y_7 &= f(y_4, y_5, y_6, x_4, x_5, x_6)\\
y_8 &= f(y_5, y_6, y_7, x_5, x_6, x_7)
\end{aligned}$$

Prof. Dr. Thomas A. Runkler

Autoregressive Models

- purely autoregressive model just considering the m previous outputs

$$y_k = \overbrace{f(y_{k-m}, \ldots, y_{k-1})}, \quad k = m + 1, \ldots, n$$

- forecast model by regression with

$$\begin{aligned}
y_4 &= f(y_1, y_2, y_3) \\
y_5 &= f(y_2, y_3, y_4) \\
y_6 &= f(y_3, y_4, y_5) \\
y_7 &= f(y_4, y_5, y_6) \\
y_8 &= f(y_5, y_6, y_7)
\end{aligned}$$

Chapter 8: Classification

1. Naive Bayes Classifier
2. Linear Discriminant Analysis
3. Support Vector Machine
4. Nearest Neighbor Classifier
5. Learning Vector Quantization
6. Decision Trees

input → continuous

Classification

output → discrete

(true/false, Healthy/sick ....)

- data set

$$Z = (\underset{\text{output}}{\overset{\text{input}}{X}}, y) = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \underset{\text{(input) continuous}}{\mathrm{IR}^p} \times \underset{\text{discrete (output)}}{\{1, \ldots, c\}}$$

- classifier

$$f : \mathrm{IR}^p \to \{1, \ldots, c\}$$

- assessment

  1. true positive (TP): $y = i,\ f(x) = i$
     (a sick patient is classified as sick)
  2. true negative (TN): $y \neq i,\ f(x) \neq i$
     (a healthy patient is classified as healthy)

errors
  3. false positive (FP): $y \neq i,\ f(x) = i$
     (a healthy patient is classified as sick) } it does not harm (false alarm)
  4. false negative (FN): $y = i,\ f(x) \neq i$
     (a sick patient is classified as healthy) } big problem (worse)

Classification Performance

- correct classifications T=TP+TN
  (number of correctly classified patients)
- false classifications F=FP+FN
  (number of incorrectly classified patients)
- relevance R=TP+FN (number of sick patients)
- irrelevance I=FP+TN (number of healthy patients)
- positivity P=TP+FP
  (number of patients that were classified as sick)
- negativity N=TN+FN
  (number of patients that were classified as healthy)
- correct classification rate T/n
  (probability that a patient is correctly classified)
- false classification rate F/n
  (probability that a patient is incorrectly classified)

Classification Performance

- true positive rate, sensitivity, recall TPR=TP/R
  (probability that a sick patient is classified as sick)
- true negative rate, specifity TNR=TN/I
  (probability that a healthy patient is classified as healthy)
- false positive rate, false alarm rate FPR=FP/I
  (probability that a healthy patient is classified as sick)
- false negative rate FNR=FN/R
  (probability that a sick patient is classified as healthy)

Classification Performance

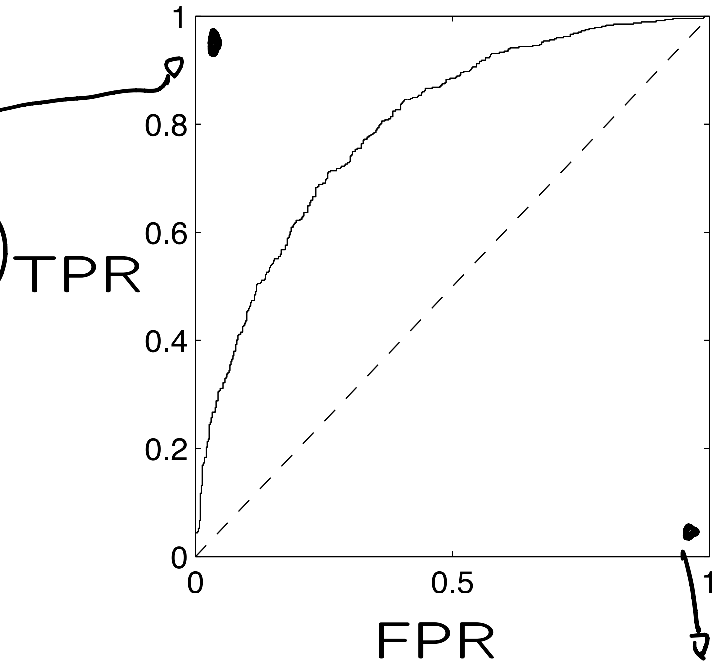
- positive prediction, precision TP/P
  (probability that a sick classified patient is sick)
- negative prediction TN/N
  (probability that a healthy classified patient is healthy)
- negative false classification rate FN/N
  (probability that a healthy classified patient is sick)
- positive false classification rate FP/P
  (probability that a sick classified patient is healthy)
- F measure $F=2/(P/TP+R/TP)=2TP/(P+R)$
  (harmonic mean of precision and recall)

Prof. Dr. Thomas A. Runkler

Classifier Diagrams

Receiver Operating Characteristic (ROC)

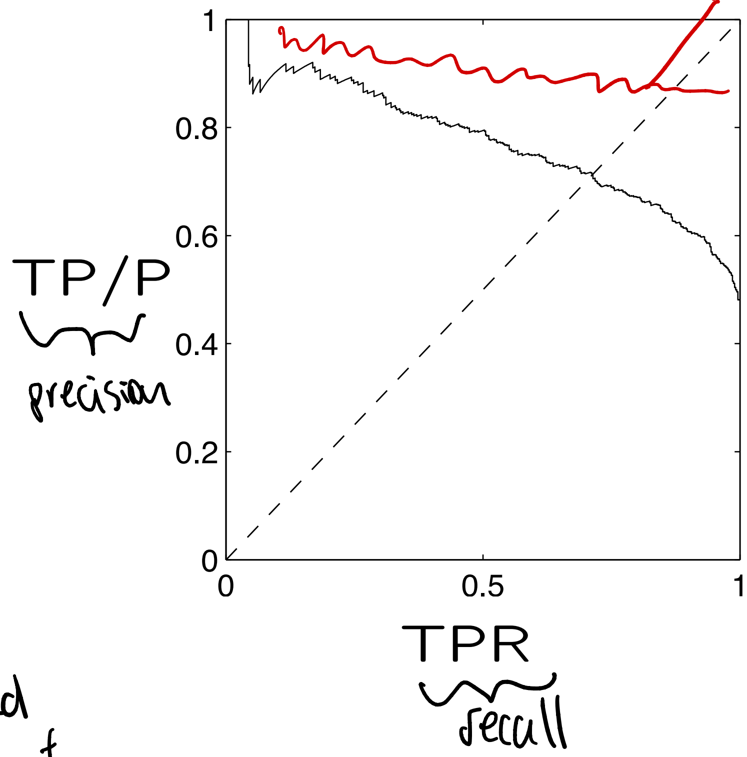
goal (the best)



very bad (but just inverting the output, it gives the best)

Precision Recall Diagram (PRD)

to stay as high as possible (the best)



probabilistic classifier

Naive Bayes Classifier

- given:
  - class probabilities

$$p(1), \ldots, p(c)$$

  - conditional feature related class probabilities

$$\begin{array}{ccc} p(x^{(1)} \mid 1), & \ldots & p(x^{(1)} \mid c) \\ \vdots & \ddots & \vdots \\ p(x^{(p)} \mid 1), & \ldots & p(x^{(p)} \mid c) \end{array}$$

- wanted: classification probabilities

$$p(1 \mid x), \ldots, p(c \mid x)$$

Naive Bayes Classifier

- naive Bayes classifier:

$$p(i \mid x) = \frac{p(i) \cdot p(x \mid i)}{\sum\limits_{j=1}^{c} p(j) \cdot p(x \mid j)}$$

→ Bayes Rule

naive assumption

(it does not always hold)

$$p(x \mid i) = \prod_{k=1}^{p} p(x^{(k)} \mid i)$$

→ individual features are independent of each other

Example Naive Bayes Classifier

	exam passed	exam failed
went to class	21	4
did not go to class	1	3
studied material	16	2
did not study material	6	5

- given: x : went to class, studied material
- wanted: $p(\text{passed} \mid x)$

Example Naive Bayes Classifier

$$p(\text{went to class}|\text{passed}) = \frac{21}{21+1} = \frac{21}{22}$$
$$p(\text{studied material}|\text{passed}) = \frac{16}{16+6} = \frac{16}{22}$$
$$\Rightarrow p(x|\text{passed}) = \frac{21 \cdot 16}{22 \cdot 22} = \frac{84}{121}$$

$$p(\text{went to class}|\text{not passed}) = \frac{4}{4+3} = \frac{4}{7}$$
$$p(\text{studied material}|\text{not passed}) = \frac{2}{2+5} = \frac{2}{7}$$
$$\Rightarrow p(x|\text{not passed}) = \frac{4 \cdot 2}{7 \cdot 7} = \frac{8}{49}$$

$$p(\text{passed}) = \frac{22}{22+7} = \frac{22}{29}$$
$$p(\text{not passed}) = \frac{7}{22+7} = \frac{7}{29}$$

$$p(\text{passed}) \cdot p(x \mid \text{passed}) = \frac{22}{29} \cdot \frac{84}{121} = \frac{168}{319}$$
$$p(\text{not passed}) \cdot p(x \mid \text{not passed}) = \frac{7}{29} \cdot \frac{8}{49} = \frac{8}{203}$$

$$\Rightarrow p(\text{passed} \mid x) = \frac{\frac{168}{319}}{\frac{168}{319} + \frac{8}{203}} = \frac{168 \cdot 203}{168 \cdot 203 + 8 \cdot 319} = \frac{147}{158} \approx 93\%$$

$+/-$ Naive Bayes Classifier

$+$ training data have to be evaluated only once

$+$ missing data can be simply ignored

$-$ features must be independent

$-$ features must be discrete

