



# The added value of machine-learning methods in forecasting wind turbine icing

---

LUKAS KUGLER

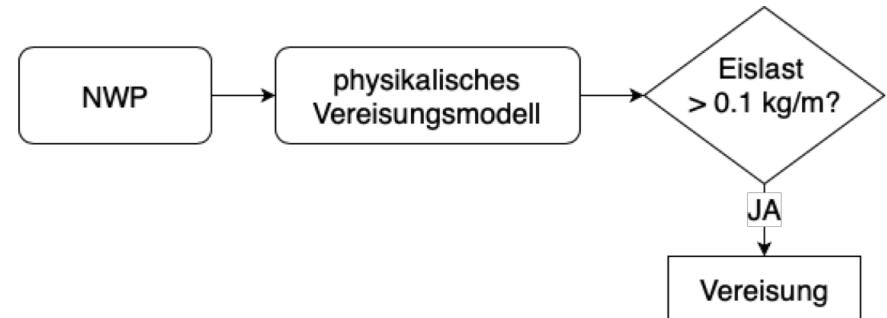
MASTERSEMINAR WS18/19

1. VORTRAG, 9.1.2019

Betreuer: Ass.-Prof. Mag. Dr. Manfred Dorninger, Dipl.-Ing. Dr. Lukas Strauss

# Vereisungsvorhersagen

Bisher meist: Kette physikalischer Modelle  
z.B.: Davis, 2014



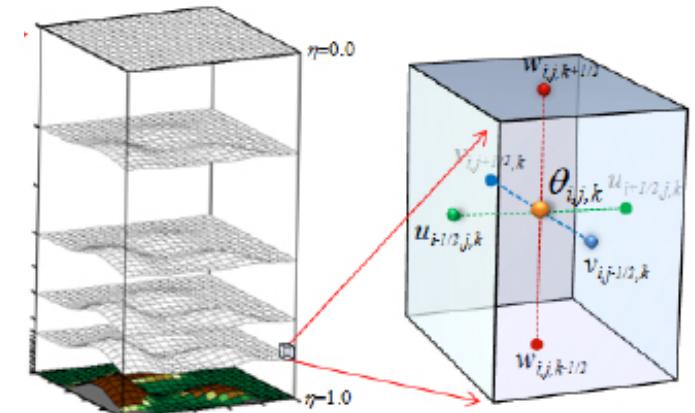
Nachteile:

„All models are wrong, but some are useful“

- Direct Model Output u.U. nicht repräsentativ
  - Parametrisierungen → Bias in T, RH
  - Wolkenwasser kaum messbar
- Vereisungsmodell
  - ursprünglich für Freileitungen konzipiert
  - abhängig von empirischen Konstanten

Erweiterung um statistische Modelle

- Nutzung von historischen und aktuellen Beobachtungsdaten
  - Berücksichtigung von modellspezifischem Bias
- Objektive Bestimmung der Vereisungskriterien (bei welcher Eislast, T, RH, zeitlicher Verlauf, ...)



©WSU

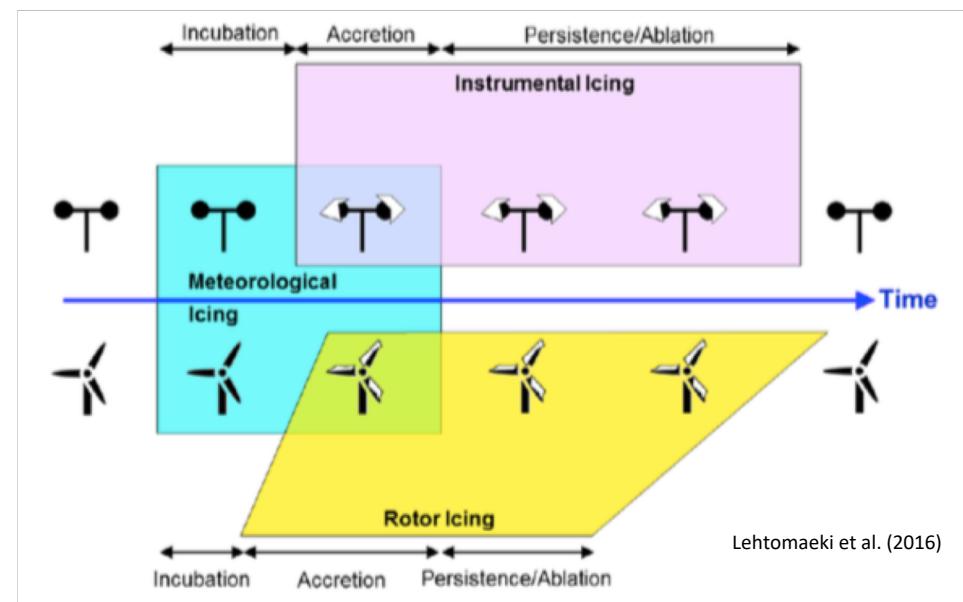
# Phasen der Vereisung

- Meteorologische Vereisung
  - Nullgradgrenze unterschritten,
  - Wolkentröpfchen vorhanden oder
  - Frostpunkt erreicht?
- Akkretion von Eis

- Instrumentelle Vereisung
  - Eis mess- oder sichtbar?
- „träges“ Vereisungsmaß

- Rotorvereisung
  - beeinflussbar durch Beheizen

→ unterschiedliche Ereignisse



# Literatur zur Vereisungsprognose

---

- Physikalische Modellierung

- Makkonen (2000): Modell für Vereisung an ‚Strukturen‘
- Oechslin (2011): Kalman gefiltertes WRF und Makkonen-Modell
- Davis (2014): NWP angetriebenes IceBlade Vereisungsmodell



- Empirisch-statistische Modellierung

- Davis (2014): Decision Tree für Vereisung und Generalized Additive Models für Power Loss
- Dabernig (2013): Prediktand Power Loss
- Sonst großteils Prognosen für Freileitungen:
  - Niu et. al. (2017): Kombinationsmodell NN-SVC-ELM
  - Zhang et. al. (2018): Detektion mit Random Forest Classifier



# Ziele und Herausforderungen

---

**Die optimale Vorhersage** berücksichtigt alle Informationsquellen:

- Historische Vorhersagen und zugehörige Beobachtungen
- Ensemble-Läufe zur Abschätzung der Unsicherheit
- Nowcasting-Information der letzten Stunde

und berücksichtigt das Cost-Loss Verhältnis des Nutzers.

## Fragestellungen

Bietet ML einen Zusatznutzen gegenüber physikalischer Modellierung?

- Welche Methode funktioniert am Besten und wieso?
- Wie hoch ist der Zusatznutzen / Unsicherheit?
- Worin liegt der Zusatznutzen?
  - Einsparung durch weniger Misses / False Alarms, mehr Hits?
- Wodurch kommt der Zusatznutzen zustande?
  - Nutzung von Ensembledaten
  - zeitliche Autokorrelation des Prediktanden (Nowcasting)
  - Nichtlineare Effekte in den Prediktoren, ...

# Daten

---

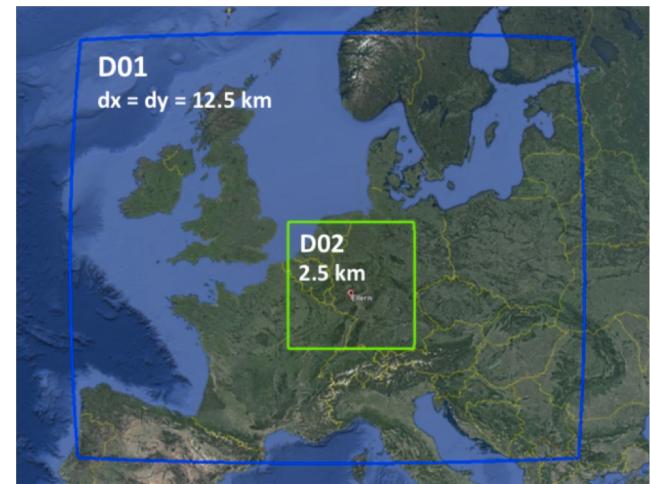
## Messdaten:

- Zeitraum: Zwei Winter 2016/17 & 2017/18
- Variablen:
  - T, RH, ff, Sichtweite, versch. Eissensoren & Kameras
  - Meteorolog. Vereisung mit Kamera manuell klassifiziert



## Vorhersagedaten:

- WRF mit 10 Ensemble Membern
- Variablen:
  - alle Modellvariablen (T, p, ff, q\_vapor, q\_cloud, ...)
- Räumliche Auflösung:
  - Bis zu 2.5 km
- Zeitliche Auflösung:
  - Stündlich, an einigen Punkten 10 minütlich

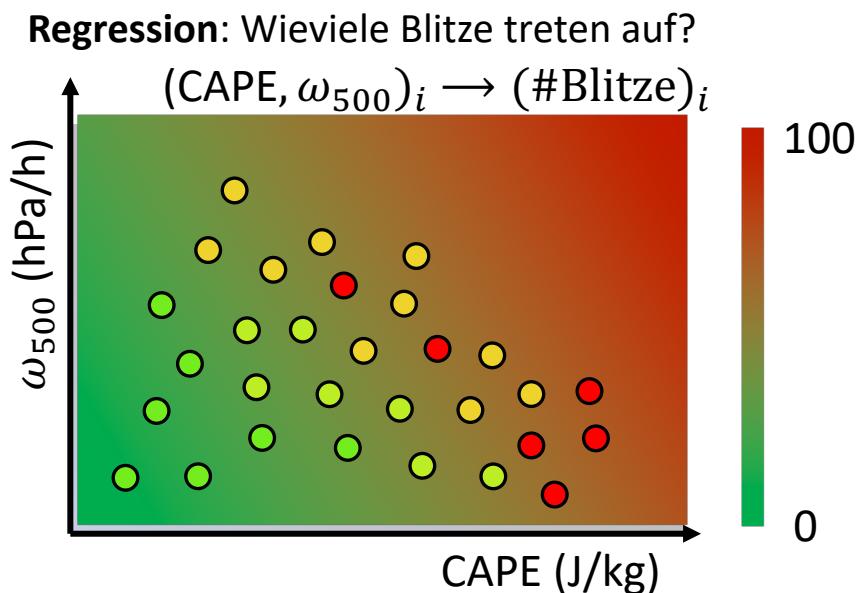
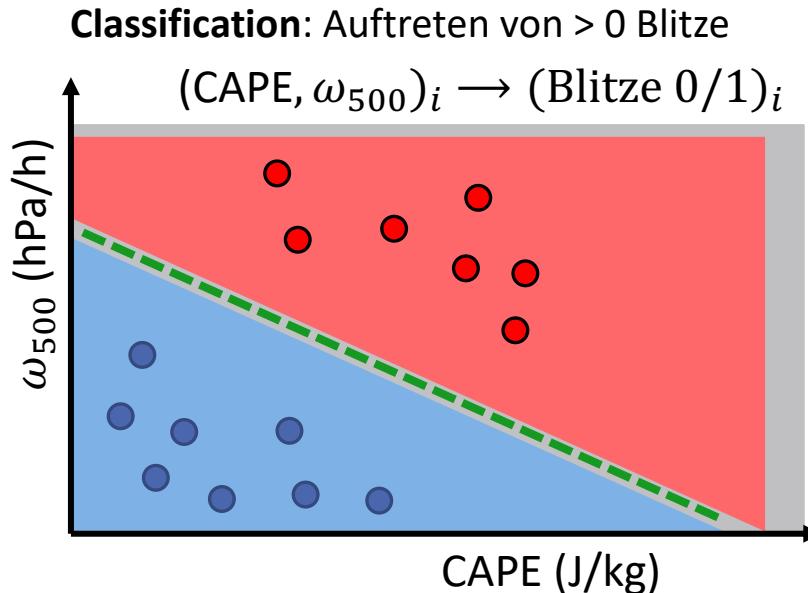


# Methoden: Machine-Learning

## Supervised Learning:

Modellierung einer Abbildung  $f: x \rightarrow y$  mit  $N$  Datenpaaren  $(x_1, y_1), \dots, (x_N, y_N)$

- wobei  $y_i$  kann je nach Methode auch vektorwertig sein, ist meist jedoch skalar, z.B.: Anzahl der Blitze
- $x_i$  typischerweise aus mehreren Prediktoren besteht, z.B.: CAPE,  $\omega_{500}$ , ...



# Methoden: Statistical Learning Theory

---

Beispiel:

Lineare Regression unter Minimierung der Fehlerquadrate

Ziel  $\tilde{y} = \alpha_0 + \alpha_1 x$

Datenpaare:  $(x_1, y_1), \dots, (x_N, y_N)$

Güte des Modells:

Summe quadrierter Fehler

$$\sum_i^N (y_i - \tilde{y}_i)^2$$

x ... Taupunkt 18z

y ... T\_min 18-06z

→ **Empirical Risk Minimization** (Vapnik & Chervonenkis)

$$L(y_i, f(x_i, \alpha)) = (y_i - f(x_i, \alpha))^2$$

$$\min \sum_i^N L(y_i, f(x_i, \alpha))$$

# Methoden: Statistical Learning Theory

Problem mit Empirical Risk Minimization?

Beispiel:

Blitzvorhersage mit *Nearest Neighbor*

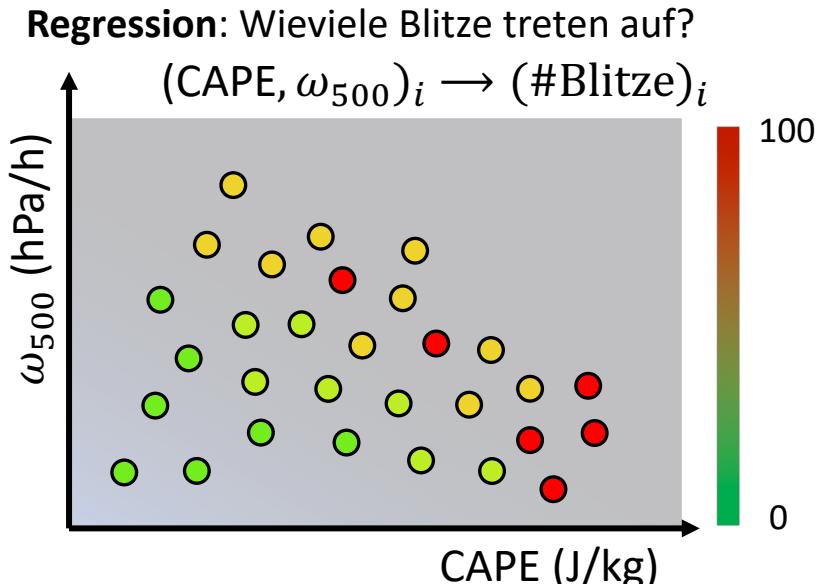
→ ~~kleiner In-Sample Error ist schlecht~~

→ nur In-Sample Error minimieren ist nicht optimal

**Gefahr von *Overfitting***

**Structural Risk Minimization** (Vapnik 1974)

→ Erwartungswert ergibt *Risk Functional*  $R(\alpha)$



P ... joint prob. distribution

$$\min \sum_i^N L(y_i, f(x_i, \alpha)) \rightarrow \min \int L(y, f(x, \alpha)) dP(x, y)$$

# Methoden: Statistical Learning Theory

---

## Beobachtungen:

- Einen kleinen In-Sample Fehler zu haben, ist nicht schwer
- Ein kleiner In-Sample Fehler heißt nicht unbedingt, dass das Modell auch gut ist

## Ergebnisse der Statistical Learning Theory (Vapnik 1974)

Was garantiert uns, dass der Fehler auch Out-of-Sample klein ist?

- Modellkomplexität (*VC-Dimension*) nicht zu hoch
- Stichprobenanzahl (iid.) nicht zu klein

d.h. wahrer Fehler ist nahe am empirischen Fehler  
Minimierung des Einen  $\Rightarrow$  Minimierung des Anderen

Aber zur numerischen Abschätzung des Out-of-Sample Fehlers zu lasch

- *Validierung* auf unabhängiger Stichprobe
- Minimierung einer Kombination aus In-Sample Error & Komplexität

$$\textcolor{red}{\textit{Regularization}} \quad J = \tilde{R}(\boldsymbol{\alpha}) + \lambda \textcolor{red}{J_P(\boldsymbol{\alpha})}$$

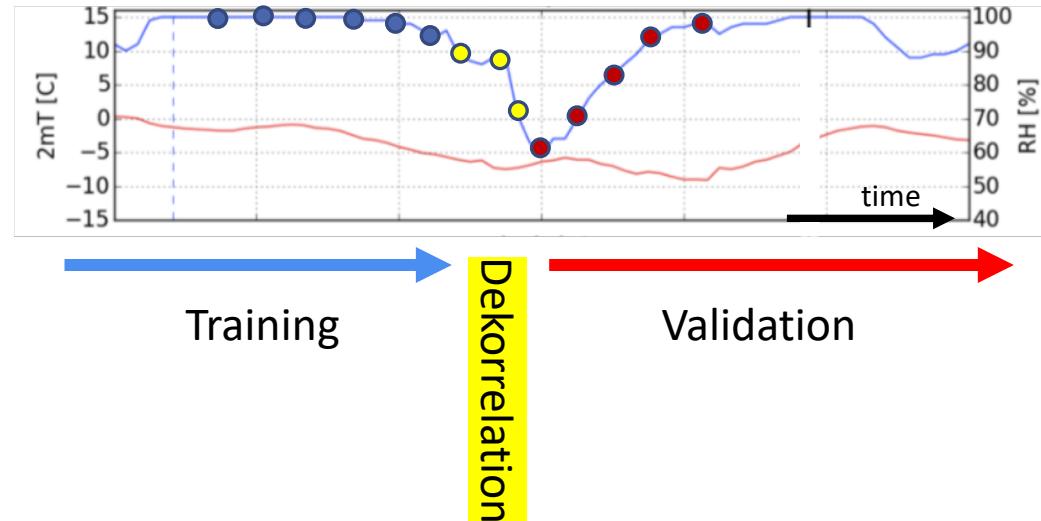
# Methoden: Validierung

**Bestimmung der Modellgüte auf einem unabhängigen Datensatz**

Meteorologische Datensätze: meist Zeitreihen (autokorriert!)

Möglichkeiten aus einer Zeitreihe zwei unabhängige Datensätze zu generieren?

- zeitliche Splits



# Methoden: Validierung

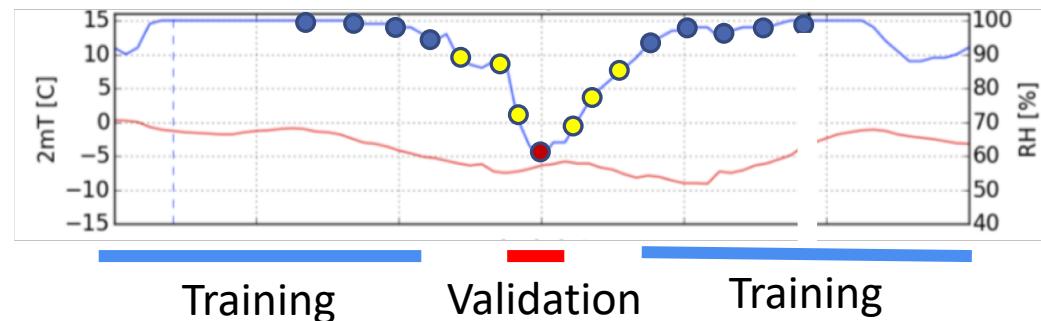
## Bestimmung der Modellgüte auf einem unabhängigen Datensatz

Meteorologische Datensätze: meist Zeitreihen (autokorriert!)

Möglichkeiten aus einer Zeitreihe zwei unabhängige Datensätze zu generieren?

- zeitliche Splits (z.B.: Training mit 2016, Validierung mit 2017)
- „h-block“ oder „hv-block CV“ (Racine, 2000)

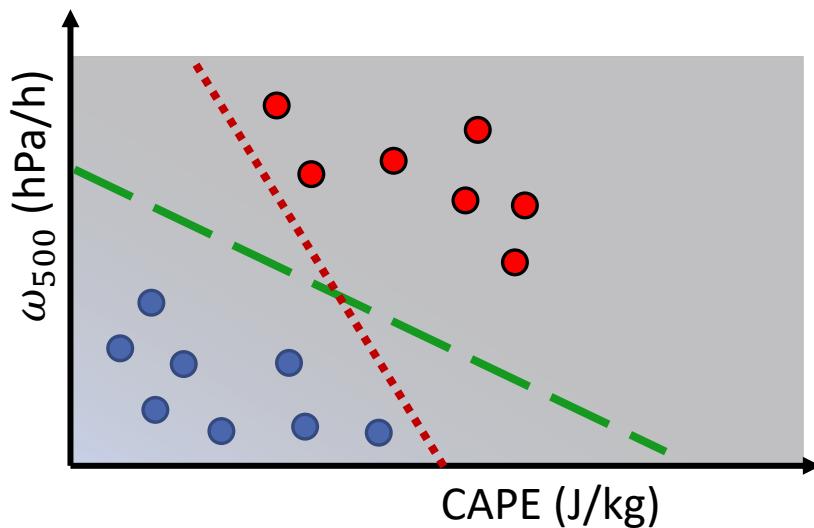
→ Datenpunkte von 6 h vor und nach dem Validierungszeitpunkt dürfen nicht ins Training eingehen



# Methoden: Support Vector Classification

**Classification:** Auftreten von > 0 Blitze

$$(\text{CAPE}, \omega_{500})_i \rightarrow (\text{Blitze } 0/1)_i$$



SVC: Beste Trennlinie, sodass Abstand zu den nächsten Punkten maximal

# Methoden: Support Vector Classification

Outcome

$$y = \begin{cases} 1 & \text{Event} \\ -1 & \text{Non-Event} \end{cases}$$

Decision Boundary  
z.B. Gerade in 2D

$$y(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + w_0 = 0$$

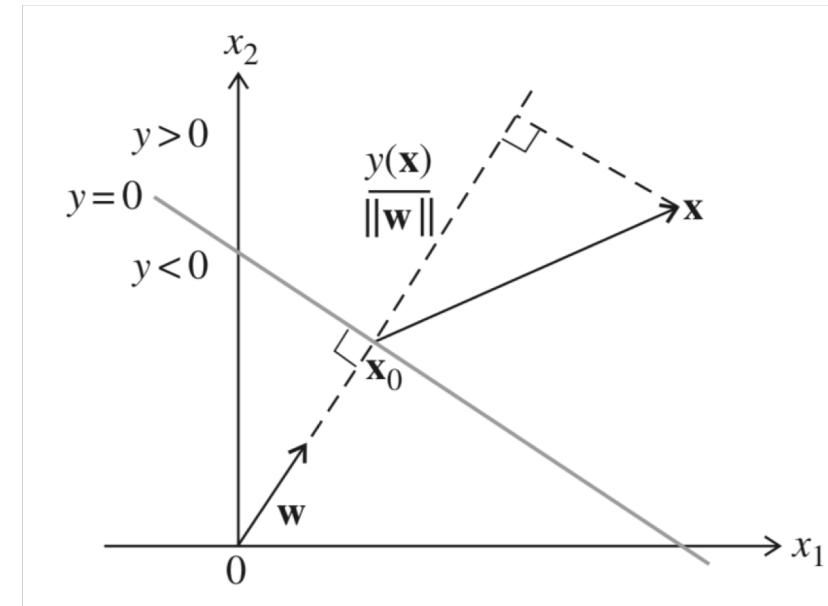
Für Punkte  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$   
auf der Geraden gilt:

$$\begin{aligned}\mathbf{w} \cdot (\mathbf{x}_1 - \mathbf{x}_2) &= 0 \\ \mathbf{w} \cdot \mathbf{x}_0 &= -w_0\end{aligned}$$

Flächeneinheitsvektor  $\hat{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$

Für einen allgemeinen  
Punkt  $\mathbf{x}$  gilt:

$$\begin{aligned}\hat{\mathbf{w}} \cdot (\mathbf{x} - \mathbf{x}_0) &= \frac{\mathbf{w} \cdot (\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{w}\|} \\ &= \frac{\mathbf{w} \cdot \mathbf{x} + w_0}{\|\mathbf{w}\|} = \frac{y(\mathbf{x})}{\|\mathbf{w}\|}\end{aligned}$$

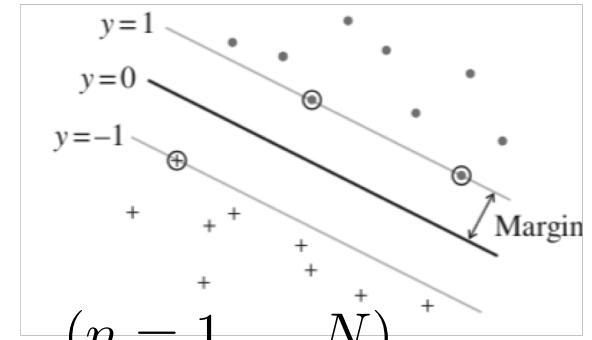


# Methoden: Support Vector Classification

Bestimmung der Trennlinie  
durch Maximierung der Margin l:

$$\max_{\mathbf{w}, w_0} l \quad \text{subject to} \quad \frac{y_{dn}(\mathbf{w} \cdot \mathbf{x}_n + w_0)}{\|\mathbf{w}\|} \geq l, \quad (n = 1, \dots, N)$$

Abstand zu den einzelnen Punkten



mit  $\|\mathbf{w}\| \equiv l^{-1}$  → ,Quadratic Programming Problem'

$$\min_{\mathbf{w}, w_0} \|\mathbf{w}\|^2 \quad y_{dn}(\mathbf{w} \cdot \mathbf{x}_n + w_0) \geq 1, \quad (n = 1, \dots, N)$$

Lösung mit Lagrange-Multiplikator-Methode

$$L_D(\boldsymbol{\lambda}) = \sum_{n=1}^N \lambda_n - 1/2 \sum_{n=1}^N \sum_{j=1}^N \lambda_n \lambda_j y_{dn} y_{dj} \mathbf{x}_n \cdot \mathbf{x}_j$$

# Methoden: Support Vector Classification

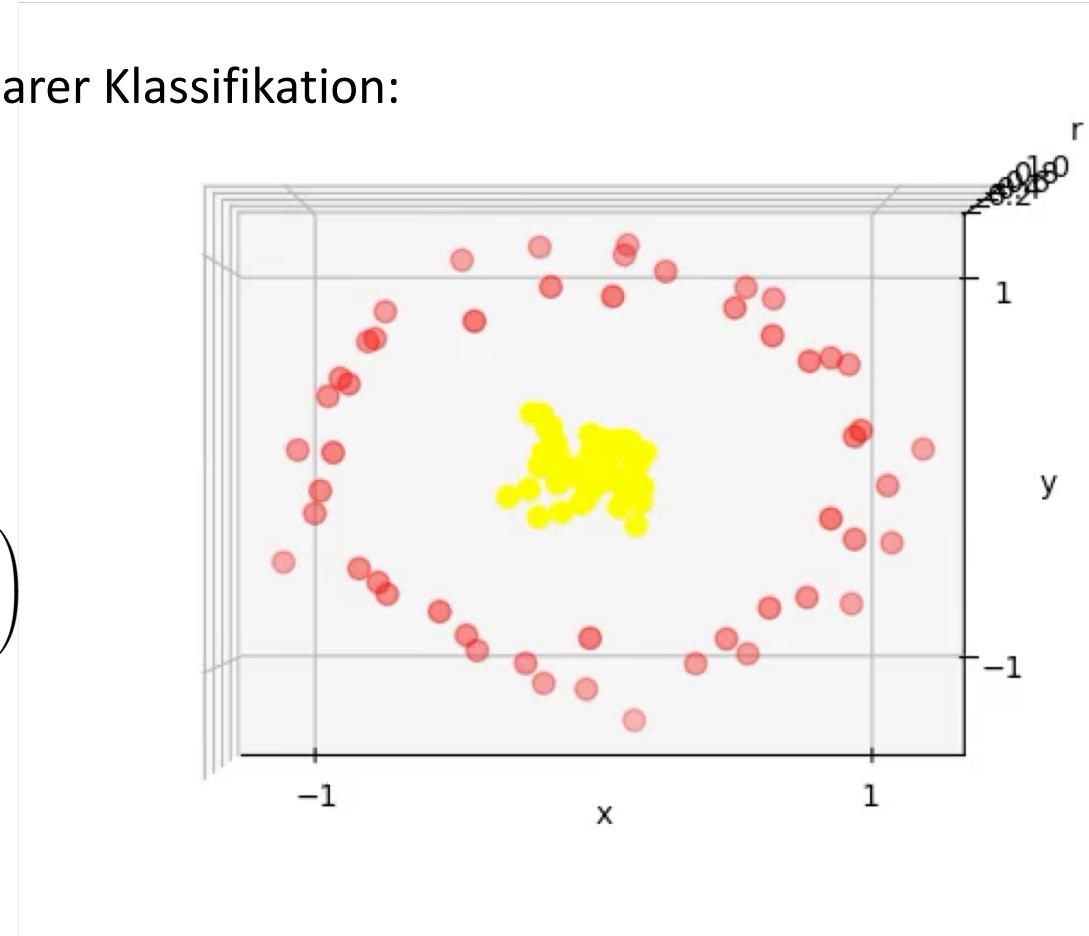
Erweiterung zu nichtlinearer Klassifikation:

$$\mathbf{x} \rightarrow \phi(\mathbf{x})$$

$$\mathbf{x}_n \cdot \mathbf{x}_j \rightarrow K(\mathbf{x}_n, \mathbf{x}_j)$$

RBF Kernel:

$$K(\mathbf{x}_n, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_n - \mathbf{x}_j\|^2}{2\sigma^2}\right)$$



# Mehrwert der Prognosen

Quantifizierung der relativen Einsparungen  
im Cost-Loss Ratio Entscheidungsmodell

Annahmen:

- Hit Rate, False Alarm Rate gegeben
- Klimatologische Häufigkeit gegeben

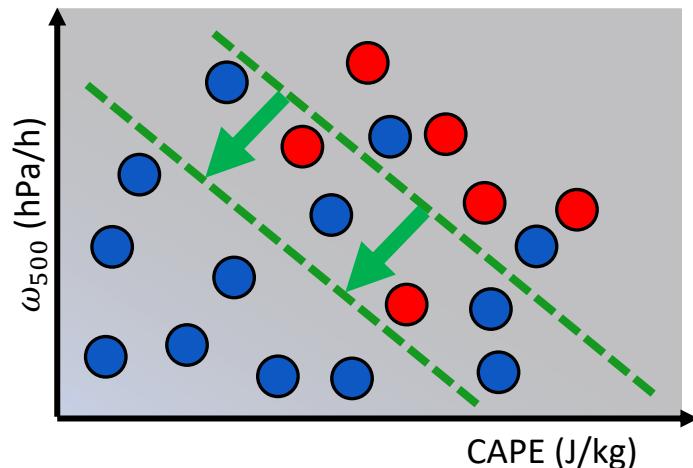
→ Einsparungen abhängig vom Nutzer (C/L)

	Yes-Event	No-Event
Yes-Action	C	C
No-Action	L	0

$$V = \frac{E_{\text{clim}} - E_{\text{forecast}}}{E_{\text{clim}} - E_{\text{perfect}}}$$

Bewertung von Wahrscheinlichkeitsprognosen

- braucht Schwellwert für binäre Prognosen
- ermöglicht Optimierung  
z.B: C/L = 0.2  $\Rightarrow$  4 False Alarms  
besser als 1 Miss
- Verschiebung des Schwellwerts  
 $\hat{=}$  Verschiebung der Decision Boundary



# The added value of ML methods in forecasting wind turbine icing

Lukas Kugler

Masterseminar WS18/19

1. Vortrag, 9.1.2019

## Einige Referenzen:

- Davis (2014), Icing Impacts on Wind Energy Production, PhD thesis
- Meteotest (2016), Final Report VGB Research Project No. 392
- Hsieh (2009), ML Methods in Environmental Sciences
- Vapnik, V. (1992). Principles of risk minimization for learning theory. In Advances in neural information processing systems (pp. 831-838) (Link)
- Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2003). Forecast verification: a practitioner's guide in atmospheric science.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning.

Diese Präsentation & weiterführende Links zu

- Publikationen zur Vereisungsvorhersage
- Einführung in Machine-Learning

auf [lkugler.github.io/mthesis](https://lkugler.github.io/mthesis)



universität  
wien



**imgw**  
Institut für Meteorologie  
und Geophysik