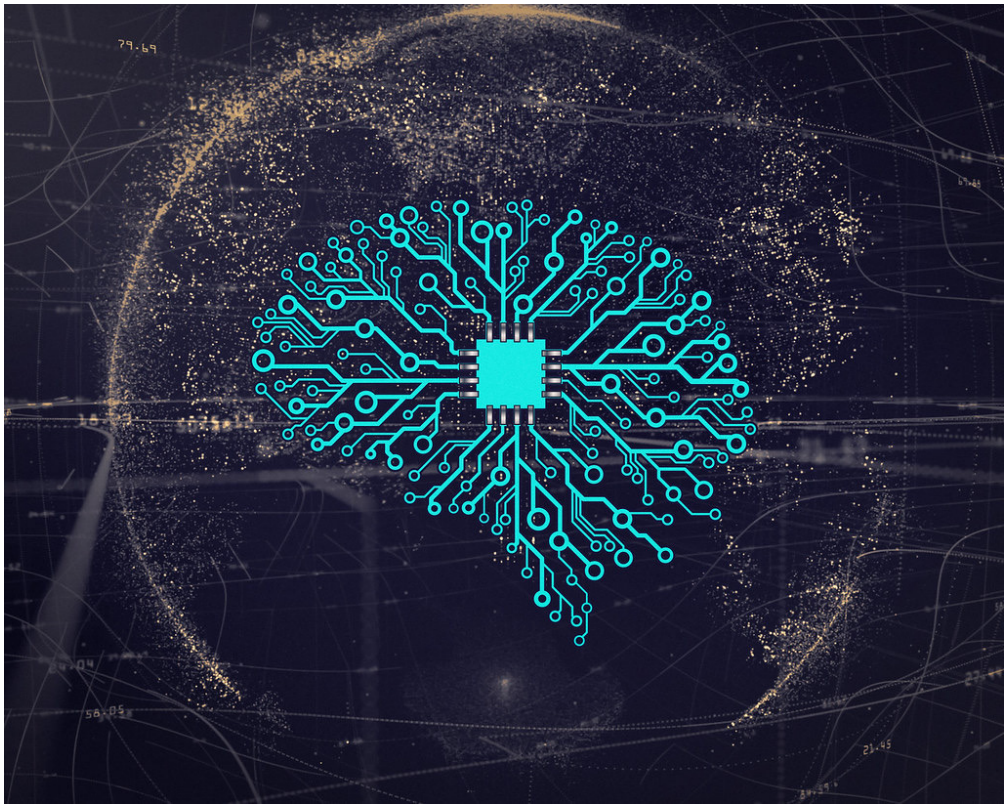
1. Einführung in die Prädiktive Modellierung

Prädiktive Modellierung - Einführung



Buzz Words:

Machine Learning, Deep Learning, Artificial Intelligence, Big Data

Exkurs: Literatur

- Pargent, F., Schoedel, R., & Stachl, C. (2023). Best Practices in Supervised Machine Learning: A Tutorial for Psychologists. Advances in Methods and Practices in Psychological Science, 6(3). https://doi.org/10.1177/25152459231162559
- James G., Witten D., Hastie T., & Tibshirani R. (2021). An Introduction to Statistical Learning (2nd edition). Springer, NY. https://www.statlearning.com/ (free book download)

Prädiktive Modellierung - Einführung

- Das Hauptziel *mancher* psychologischen Fragestellungen/Anwendungen besteht darin, möglichst präzise Vorhersagen einer interessierenden Variable (Kriterium), basierend auf einer Reihe von Prädiktorvariablen zu treffen*

- **Beispiel 1:**
  - In der **Personalpsychologischen Eignungsdiagnostik** ist es das Ziel der Unternehmen, möglichst passende Bewerber*innen zu finden
  - Vorhersage von Berufserfolg mithilfe von Variablen aus dem Bewerbungsverfahren (IQ-Test, Persönlichkeitstest, Aufgaben im Assessmentcenter)

*Ein anderes Hauptziel psychologischer Forschung besteht darin, *kausale* Zusammenhänge zu identifizieren.

Prädiktive Modellierung - Einführung

- Das Hauptziel *mancher* psychologischen Fragestellungen/Anwendungen besteht darin, möglichst präzise Vorhersagen einer interessierenden Variable (Kriterium), basierend auf einer Reihe von Prädiktorvariablen zu treffen

- **Beispiel 2:**
  - In der **klinischen Psychologie** ist es ein Ziel Suizidversuche zu verhindern
  - Vorhersage von Suizidversuchen durch Smartphone-Nutzung/Social-Media Verhalten

Prädiktive Modellierung - Einführung

- Das Hauptziel *mancher* psychologischen Fragestellungen/Anwendungen besteht darin, möglichst präzise Vorhersagen einer interessierenden Variable (Kriterium), basierend auf einer Reihe von Prädiktorvariablen zu treffen

- **Beispiel 3:**
  - In der **Verkehrspsychologie** möchte man sicher stellen, dass Personen nur in fahrtauglichem Zustand Auto fahren
  - Vorhersage von Fahrtauglichkeit mithilfe von Kameradaten (Lidschlag etc.), chemischen Daten (Alkoholmessgerät im Auto), Lenkverhalten (Mikrobewegungen etc.)

Prädiktive Modellierung - Einführung

- Prädiktive Modellierung stellt Methoden bereit, um möglichst präzise statistische Vorhersagemodelle zu erstellen.

Agenda für heute

- Einführung zum prädiktiven Modell
  - Definition
  - Modellklassen
  - Annahmen
  - Regression vs. Klassifikation
  - Flexibilität & Interpretierbarkeit
- Quantifizierung der Vorhersagegüte
- Bias-Varianz Tradeoff

Agenda für heute

- **Einführung zum prädiktiven Modell**
  - Definition
  - Modellklassen
  - Annahmen
  - Regression vs. Klassifikation
  - Flexibilität & Interpretierbarkeit
- Quantifizierung der Vorhersagegüte
- Bias-Varianz Tradeoff

Prädiktives Modell

- **Definition:** Ein *prädiktives Modell* ist ein Algorithmus, der für eine beliebige Beobachtung i bei gegebenen Prädiktorwerten x_{i1} bis x_{ip} eine konkrete Vorhersage $\hat{y}_i$ für die Kriteriumsvariable Y liefert
- Beispiel: Prädiktion von Berufserfolg
  - Prädiktorwerte für eine konkrete Person j sind:
    - $x_{j,IQfluid} = 105$
    - $x_{j,IQkristal} = 115$
    - $x_{j,Neuro} = 110$
    - $x_{j,Extra} = 80$
    - $x_{j,Offen} = 85$
    - $x_{j,Vert} = 108$
    - $x_{j,Gewis} = 120$
  - Prädiktives Modell liefert als Vorhersage: $\hat{y}_{j,Berufserfolg} = 106$

Prädiktives Modell: Trainingsphase

- Verschiedene Modellklassen besitzen unterschiedliche Modellparameter.
- Bevor konkrete Vorhersagen berechnet werden können, müssen die Modellparameter geschätzt werden. Man sagt auch, das prädiktive Modell wird „trainiert“.
- Wird für die Modellschätzung ein Datensatz verwendet, in dem sowohl die Werte der Prädiktoren als auch die Werte der Kriteriumsvariable bekannt sind, spricht man auch von „**Supervised (Machine) Learning**“.

Beispiel für eine prädiktive Modellklasse

- Modellklasse: multiple lineare Regression
  - $Y_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i$
- Modellparameter: α, β_1, ... , β_p
  - Schätzung durch Methode der kleinsten Quadrate anhand der vollständigen Daten von Beobachtungen i = 1, ... , N
  - Liefert die Schätzwerte $a, b_1, \ldots, b_p$
- Vorhersage für neue Beobachtung j (y_j unbekannt):
  - $\hat{y}_j = a + b_1 x_{j1} + \cdots + b_p x_{jp}$

Festes vs. Zufälliges Modell

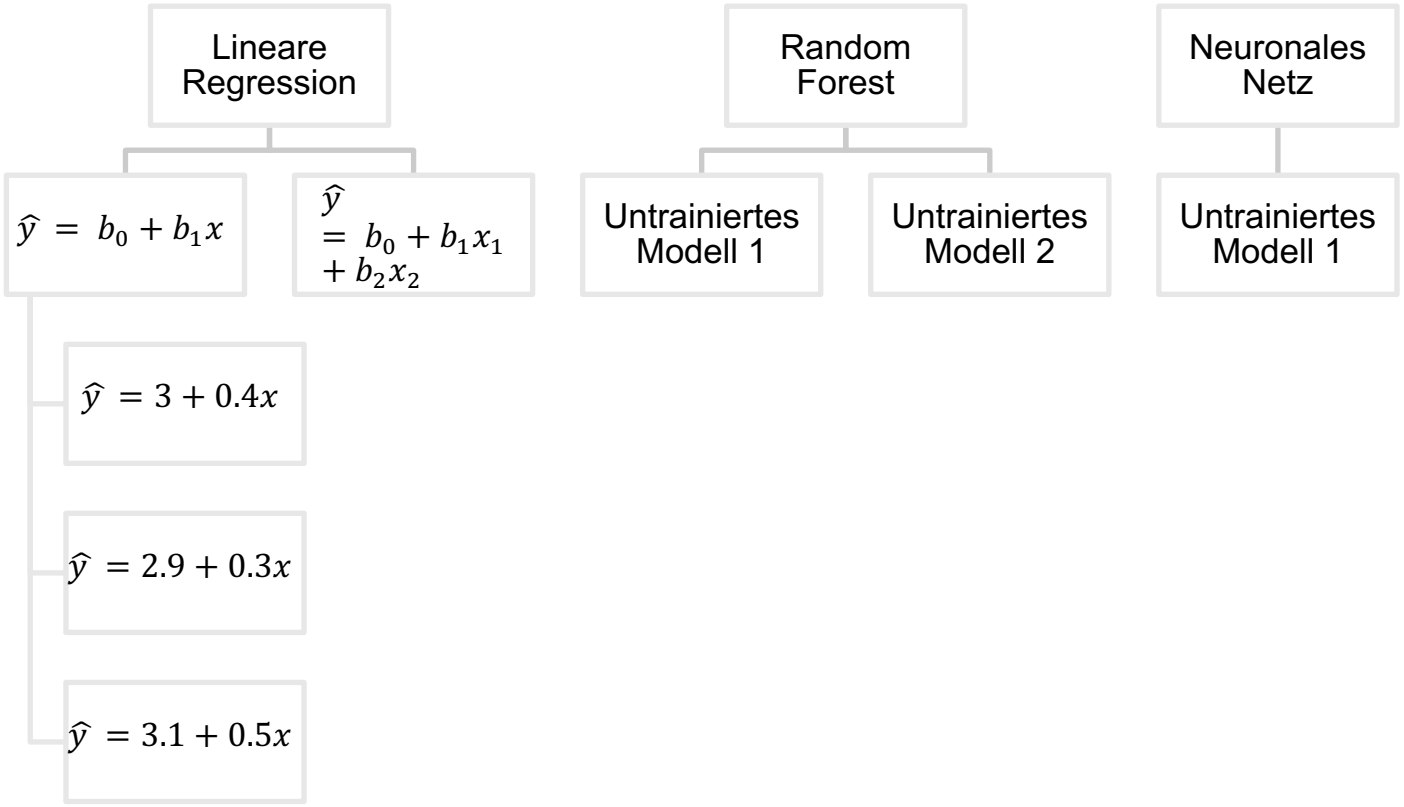
- Manchmal ist es nützlich, zwischen einem **trainierten (festen) Modell** und dem **noch nicht trainierten (zufälligen) Modell** zu unterscheiden.
- Dabei meinen wir *zufällig* im Sinne einer Zufallsvariable
- Ein Modell ist zufällig, solange die Modellparameter noch nicht anhand einer konkreten Stichprobe geschätzt wurden. Damit können mit dem zufälligen Modell noch keine konkreten Vorhersagen getroffen werden.
- Bei einem zufälligen Modell stehen folgende Informationen bereits fest:
  - Population (z.B. Psychologiestudierende in Deutschland)
  - Kriterium (z.B. Extraversion)
  - Prädiktoren (z.B. Alter, Abiturnote, ...)
  - Modellklasse (z.B.: multiple lineare Regression)
- In anderen Worten: bei einem zufälligen/nicht trainierten Modell steht zwar die Model*struktur* schon fest, die genauen numerischen Werte der Modellparameter aber noch nicht.

Modellklasse – zufällige Modell – festes (trainiertes) Modell

Modellklasse

Zufällige Modelle
(= konkrete Modellstruktur, untrainierte Parameter)

Feste Modelle
(= an verschiedenen Stichproben trainierte Modelle mit fixierten Parametern)



Eigenschaften der Beobachtungen

- In der prädiktiven Modellierung gehen wir in der Regel davon aus, dass alle Beobachtungen, für die mit einem trainierten prädiktiven Modell Vorhersagen getroffen werden sollen, aus der gleichen Population stammen wie die Beobachtungen, die zum Trainieren des Modells verwendet wurden.
- Alle Beobachtungen sollen außerdem unabhängige Ziehungen aus dieser einen Population sein.
- Beides impliziert, dass sich die Zusammenhänge in der Population nicht verändert haben, seit das Modell trainiert wurde.
- Nur mit diesen Annahmen ist es möglich, für neue Beobachtungen die Genauigkeit von Vorhersagen statistisch sinnvoll abzuschätzen.
- Technisch ist es natürlich ohne weiteres möglich, Vorhersagen für Beobachtungen aus einer anderen Population zu berechnen. Um die Genauigkeit solcher Vorhersagen abzuschätzen, muss das Modell für diese Population aber neu evaluiert werden.

Flexibilität

- Verschiedene Modellklassen unterscheiden sich in ihrer Flexibilität
- Bsp.: multiple lineare Regression → relativ niedrige Flexibilität
  - Nur lineare Zusammenhänge zwischen Prädiktoren und Vorhersagen
  - Keine Interaktionen zwischen Prädiktoren
  - Nur wenige Prädiktoren möglich, sonst Modell oft instabil
- Ist der wahre Zusammenhang nonlinear, liegen Interaktionen vor, oder hängen sehr viele Variablen mit dem Kriterium zusammen, kann eine flexiblere Modellklasse eventuell präzisere Vorhersagen erzielen.
- Modellklassen die Nonlinearität, Interaktionen, viele Prädiktoren berücksichtigen können: Random Forests, Gradient Boosting, Support Vector Machines, neurale Netze („Deep Learning“)
- Wie werden in dieser Vorlesung v.a. den Random Forest behandeln.

Interpretierbarkeit

- **Definition:** Ein prädiktives Modell ist interpretierbar, wenn man versteht, wie die Werte in den Prädiktorvariablen mit der getroffenen Vorhersage zusammen hängen (keine „black box“).
  - z.B.: höherer vorhergesagter Berufserfolg für Personen mit höheren Werten in Extraversion
    ($b_{Extraversion} > 0$ in multipler linearer Regression)
- Oft liegt ein Tradeoff zwischen Modellflexibilität und Interpretierbarkeit vor. Für viele flexible Modellklassen ist Interpretierbarkeit nicht direkt gegeben und kann nur durch Umwege erzielt werden.
- In der prädiktiven Modellierung ist unser Hauptziel, möglichst präzise Vorhersagen zu treffen. Damit sind wir bereit, nicht interpretierbare Modelle zu verwenden, sofern wir nachweisen können, dass sie für unseren konkreten Anwendungsfall präzisere Vorhersagen erzielen.

Regression vs. Klassifikation

Unterscheidung zweier Prädiktionsszenarien:

- **Regression**:
  - Kriteriumsvariable *kontinuierlich*
    z.B. Alter, Depressivität, Intelligenz
- **Klassifikation**:
  - Kriteriumsvariable besteht aus *Kategorien*
    z.B. Studienerfolg (Ja/Nein), Suizidversuch (Ja/Nein), Psychische Störungen
  - Vorlesung beschränkt sich auf binäre Klassifikation
    (d.h. Kriteriumsvariable kann nur zwei verschiedene Werte annehmen)

- Je nach Szenario unterscheiden sich:
  - Modellklassen (z.B. lineare Regression vs. logistische Regression)
  - Evaluation der Vorhersagegüte (z.B. MSE vs. MMCE; siehe unten)

Agenda für heute

- Einführung zum prädiktiven Modell
  - Definition
  - Modellklassen
  - Annahmen
  - Regression vs. Klassifikation
  - Flexibilität & Interpretierbarkeit
- **Quantifizierung der Vorhersagegüte**
- Bias-Varianz Tradeoff

Performancemaße: Quantifizierung der Vorhersagegüte

- Wie quantifiziert man die Genauigkeit von Vorhersagen bzw. wie quantifiziert man den Vorhersagefehler?
- Wir nehmen dafür zunächst an, es liegen für eine Reihe von Beobachtungen nur die tatsächlichen Werte y_i sowie die Vorhersagen $\hat{y}_i$ vor. Woher die Vorhersagen kommen ist nicht bekannt, bzw. irrelevant.
  - z.B.:

Proband	y	$\hat{y}$
1	102	105
2	105	103
3	109	110
4	98	100
5	110	108
6	95	99

- Als Kriterien für die Präzision von Vorhersagen werden sogenannte „Performancemaße“ verwendet.
- Wahl der Performancemaße ist abhängig vom Prädiktionssetting (Regression oder Klassifikation?)
- Es gibt Standardmaße oder maßgeschneiderte Maße

Performancemaße Regression

- Mean Squared Error:
  - $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
  - N: Anzahl von Beobachtungen

- Determinationskoeffizient:
  - $R^2 = 1 - \frac{qs_{res}}{qs_{ges}} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$

- Weitere Maße: Root Mean Squared Error, Median Absolute Error, ...

Performancemaße Klassifikation

- Mean Misclassification Error
  - $MMCE = \frac{1}{N}\sum_{i=1}^{N} I(y_i \neq \hat{y}_i)$
  - $I(y_i \neq \hat{y}_i) = \begin{cases} 1, y_i \neq \hat{y}_i \\ 0, y_i = \hat{y}_i \end{cases}$, I heißt Indikatorfunktion

- MMCE entspricht dem Anteil falsch klassifizierter Fälle.

- In der Regel werden die beiden Ausprägungen einer binären Kriteriumsvariable mit den Werten 0 und 1 kodiert (Dummy-Kodierung).
- Bei binärer Klassifikation mit $y_i \in \{0,1\}$ entspricht der MMCE dem MSE.

Performancemaße Klassifikation

- MMCE ist oft nicht ausreichend, um die Güte von Vorhersagen angemessen zu beurteilen.
- Im Klassifikationsfall können Vorhersagen und tatsächliche Kriteriumswerte in einer sogenannten Konfusionsmatrix dargestellt werden.
- Beispiel:

A)

$\hat{y}_i$ \ y_i	1	0
1	400	200
0	0	200

B)

$\hat{y}_i$ \ y_i	1	0
1	300	100
0	100	300

- In beiden Fällen ist $MMCE = 0{,}25$ ($MMCE_A = \frac{0+200}{800}$, $MMCE_B = \frac{100+100}{800}$).
- Inhaltlich unterscheiden sich die beiden Konfusionsmatrizen jedoch deutlich voneinander.
- Wenn möglich, immer die komplette Konfusionsmatrix betrachten.
- Performancebeurteilung bei Klassifikation ist daher meistens schwieriger als bei Regression.

Performancemaße Klassifikation

- TP: Anzahl richtig klassifizierter Merkmalsträger*innen („True Positives“)
- TN: Anzahl richtig klassifizierter Nicht-Merkmalsträger*innen („True Negatives“)
- FP: Anzahl fälschlicherweise als Merkmalsträger*innen klassifizierte Nicht-Merkmalsträger*innen („False Positives“)
- FN: Anzahl fälschlicherweise als Nicht-Merkmalsträger*innen klassifizierte Merkmalsträger*innen („False Negatives“)
- Sensitivität: $SENS = \frac{TP}{TP+FN}$
- Spezifität: $SPEC = \frac{TN}{TN+FP}$
- Positiver Prädiktionswert: $PPV = \frac{TP}{TP+FP}$
- Negativer Prädiktionswert: $NPV = \frac{TN}{TN+FN}$
- Genauigkeit: $ACC = \frac{TP+TN}{N} = 1 - \frac{FP+FN}{N} = 1 - MMCE$

$\hat{y}_i$ \ y_i	1	0
1	TP	FP
0	FN	TN

$$TP + FP + FN + TN = N$$

Performancemaße Klassifikation

- MMCE kann nur sinnvoll interpretiert werden, wenn die Prävalenz $PREV = \frac{TP+FN}{N}$ bekannt ist.
- Bei extremen Prävalenzen kann der MMCE stark in die Irre führen.
- Beispiel:

C)

$\hat{y}_i$ \ y_i	1	0
1	20	20
0	80	880

$$PREV = \frac{TP + FN}{N} = 0{,}1$$

$$MMCE = \frac{FP + FN}{N} = 0{,}1$$

$$SENS = \frac{TP}{TP + FN} = 0{,}2$$

- **Problem:** Für ein triviales prädiktives Modell, das für jede Beobachtung i automatisch $\hat{y}_i = 0$ vorhersagt gilt: $MMCE = PREV = 0{,}1$

Performancemaße Klassifikation

- Je nach Anwendungsfall sind manche Performancemaße von größerer praktischer Bedeutung als andere.
- Beispiele:
  - Personalauswahl: Anteil geeigneter Bewerber*innen an den Eingestellten relevant
    → Positiver Prädiktionswert
  - Klinische Diagnosen: Anteil richtig erkannter Erkrankten an den Kranken
    → Sensitivität
- **Problem:** Das triviale prädiktive Modell, welches für jede Beobachtung i automatisch $\hat{y}_i = 1$ vorhersagt, hat immer eine Sensitivität von 1 (äquivalent auch für Spezifität möglich).
- Man betrachtet daher oft Sensitivität und Spezifität gemeinsam
- Weitere Maße: Area Under the Curve, Cohen‘s Kappa, ...

Agenda für heute

- Einführung zum prädiktiven Modell
  - Definition
  - Modellklassen
  - Annahmen
  - Regression vs. Klassifikation
  - Flexibilität & Interpretierbarkeit
- Quantifizierung der Vorhersagegüte
- **Bias-Varianz Tradeoff**

Gedankenexperiment

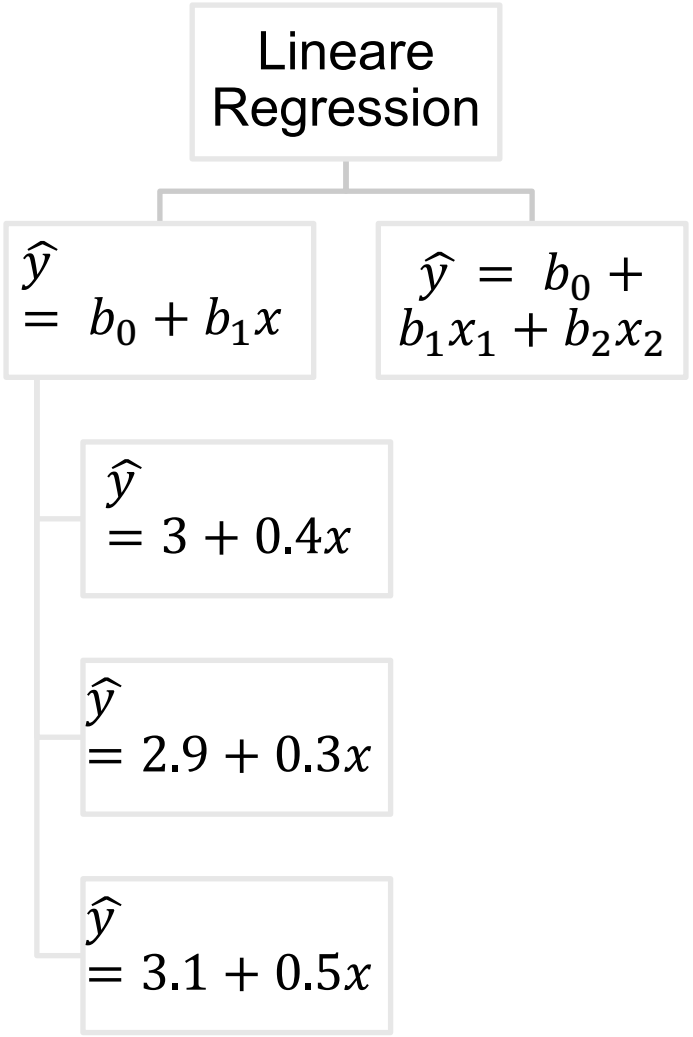
- Uns liegen zwei trainierte prädiktive Modelle vor, sowie die folgenden Informationen (identisch für beide Modelle):
  - Kriterium, welches durch die Modelle vorhergesagt werden kann
  - Notwendige Prädiktoren, um eine Vorhersage berechnen zu können
  - Population, für welche die Modelle erstellt wurden
- Als Anwender*innen interessiert uns, wie gut die beiden Modelle funktionieren: Wie genau sind Vorhersagen für beliebige Individuen aus der Population?
- Idealisiertes Vorgehen:
  - Erhebe eine sehr große Stichprobe aus der Population
  - Berechne Vorhersagen mithilfe der beobachteten Prädiktorwerte
  - Vergleiche die Vorhersagen mit den ebenfalls erhobenen tatsächlichen Kriteriumswerten, durch die Berechnung eines geeigneten Performancemaßes

Zusammensetzung des Vorhersagefehlers

- Warum unterscheiden sich trainierte prädiktive Modelle hinsichtlich der Güte ihrer Vorhersagen?
- Der erwartete Vorhersagefehler eines zufälligen prädiktiven Modells hängt von den folgenden drei Größen ab:
  - **Bias**
  - **Varianz**
  - **Nicht reduzierbarer Fehler**
- Je höher eine der Größen, desto höher der erwartete Vorhersagefehler (unter Annahme der Konstanthaltung der anderen beiden Größen)
- Da wir den wahren Zusammenhang in der Population niemals kennen, kann keine der drei Größen in der Praxis exakt bestimmt werden.
- Trotzdem sind diese Größen als Metaheuristiken zum Verständnis von prädiktiver Modellierung extrem hilfreich.

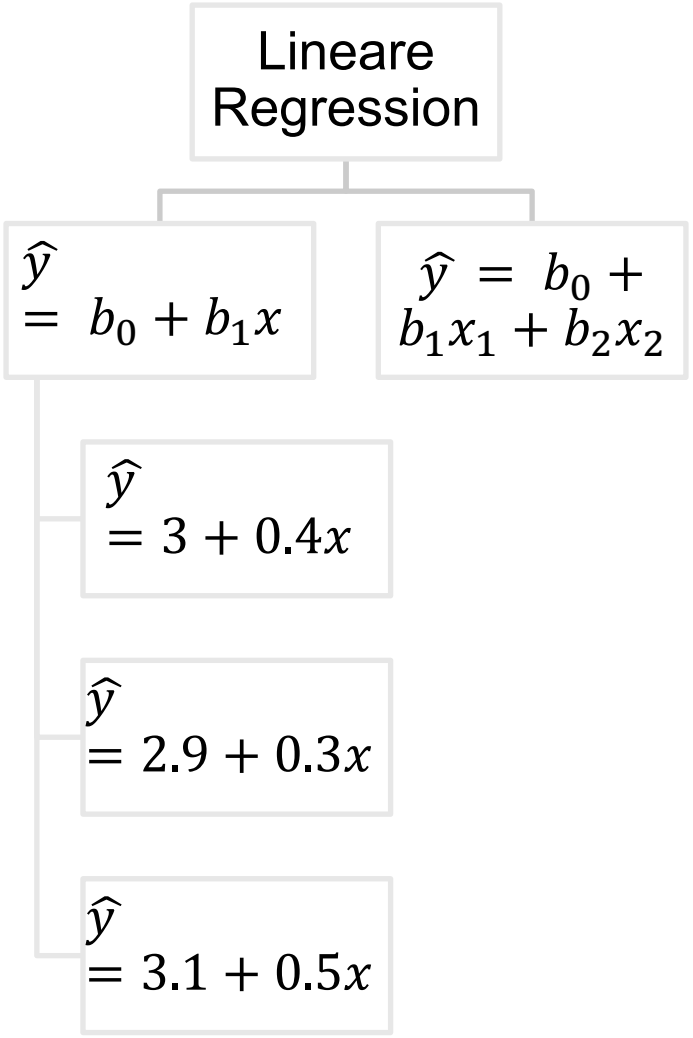
Bias

- Es liegt *kein* Bias vor: Für alle Beobachtungen sagt ein zufälliges Modell im Mittel (d.h., über viele konkret trainierte feste Modelle) den richtigen erwarteten Wert vorher
- Es liegt Bias vor: das zufällige Modell sagt für manche Beobachtungen im Mittel nicht den richtigen erwarteten Wert vorher
- Selbst wenn flexible Modellklassen theoretisch den wahren Zusammenhang abbilden können, müssen Modellparameter immer anhand von Stichprobendaten geschätzt werden.
- D.h., jedes konkrete trainierte Modell wird mal daneben liegen – die Frage beim Bias ist, ob alle (hypothetischen) trainierten Modelle eines zufälligen Modells *im Mittel* den richtigen Wert vorhersagen
- Ein niedriger Bias allein ist für einen niedrigen erwarteten Vorhersagefehler aber nicht ausreichend…

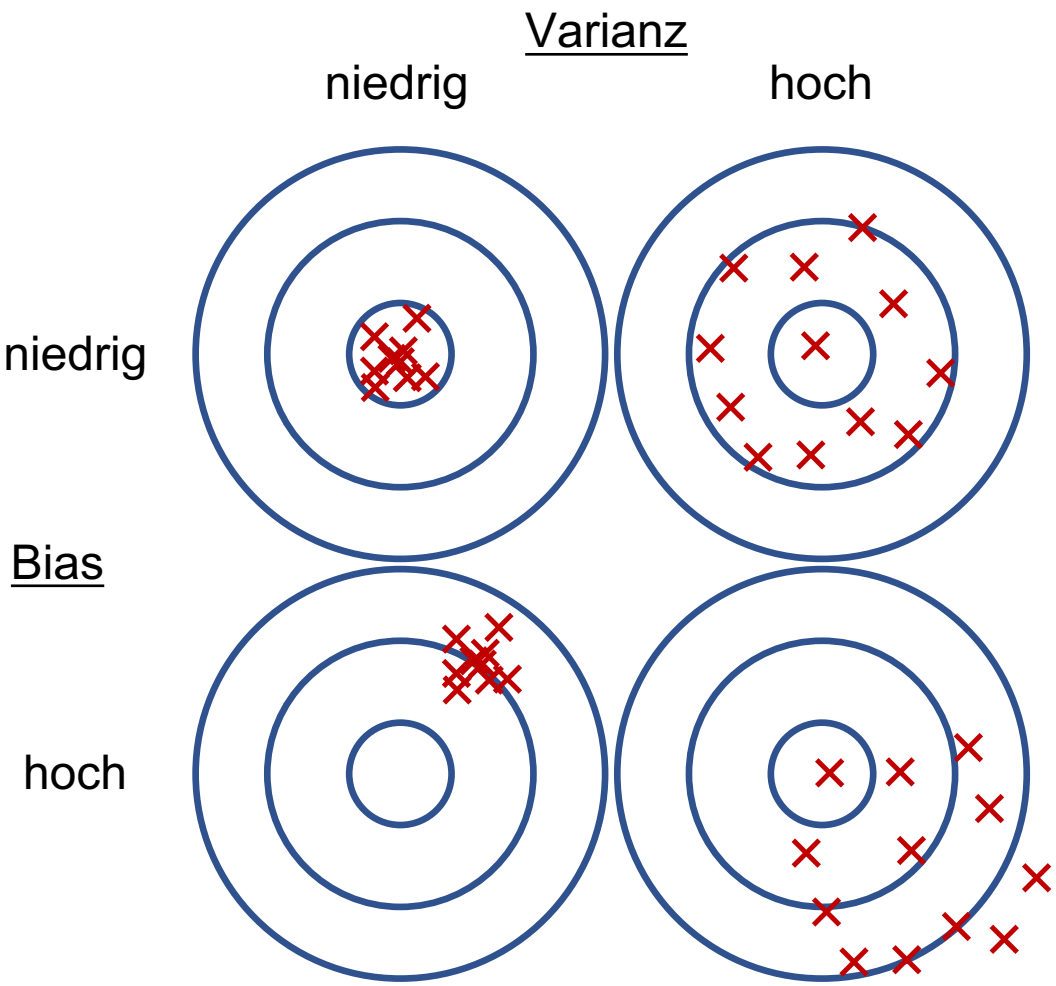


Varianz

- Wie stabil ist ein Modell bzw. wie stabil sind dessen Vorhersagen?
- Wie stark unterscheiden sich die Vorhersagen für beliebige feste Beobachtungen, wenn das zufällige Modell wiederholt mit Stichproben der gleichen Größe trainiert wird?
- Beispiel:
  - Ziehe 1000 Stichproben und trainiere jeweils eine multiple lineare Regression
  - Ziehe eine weitere Beobachtung und vergleiche die Vorhersagen der 1000 trainierten Modelle
- Je flexibler die Modellklasse, desto größer ist tendenziell die Varianz
- Je größer die ursprünglichen Stichproben, desto geringer ist die Varianz (siehe z.B. Standardfehler in der multiplen linearen Regression)



Zielscheiben Metapher



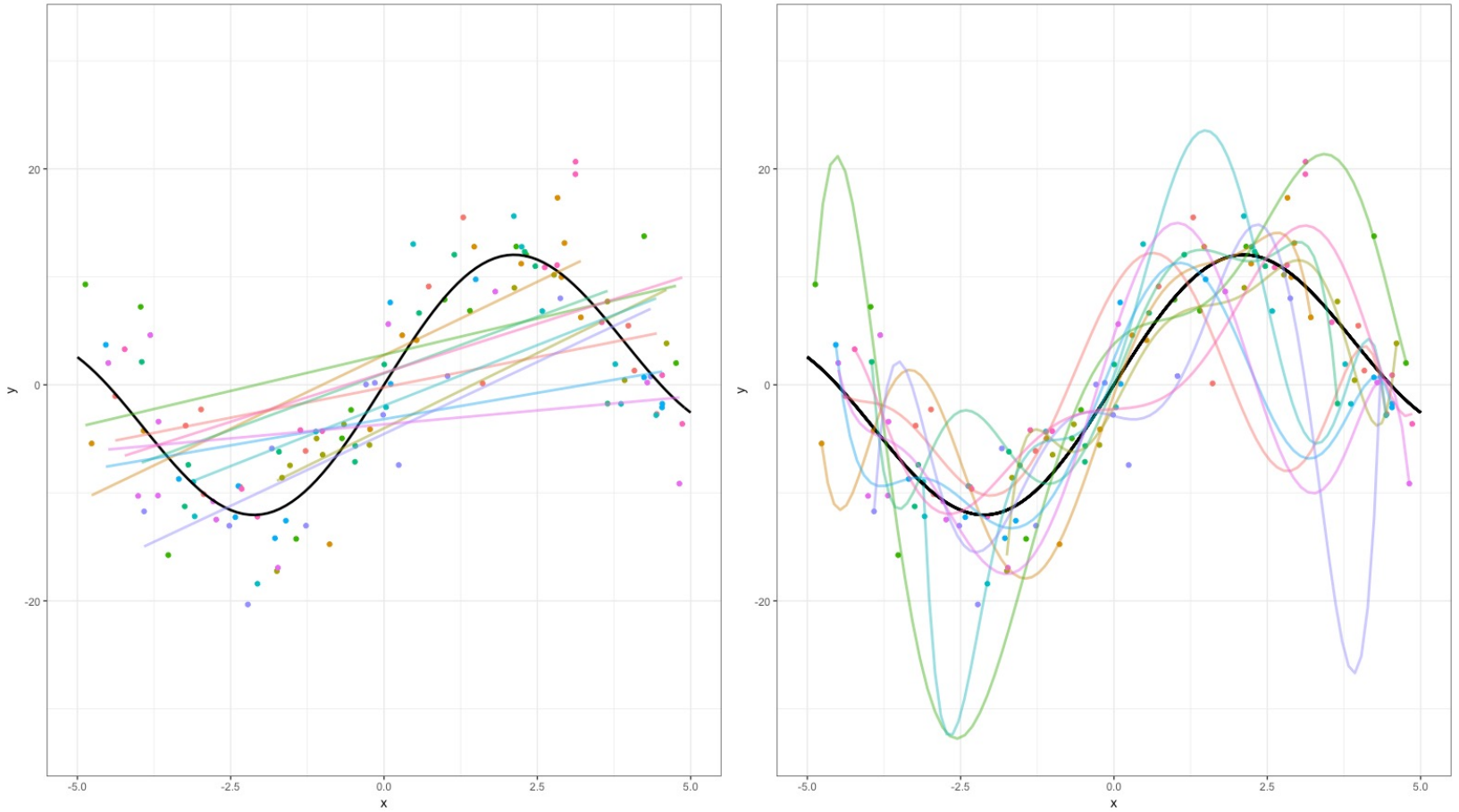
Nicht reduzierbarer Fehler

- Stochastischer Fehler, der auch in der Population vorliegt
- Resultiert nicht durch die Schätzung des Modells bei endlichen Stichproben
- Liegt kein nicht reduzierbarer Fehler vor, kann das Kriterium mithilfe der Prädiktoren theoretisch deterministisch vorhergesagt werden.
  - Praktisch nur in trivialen Anwendungen möglich
  - Unmöglich, sobald Prädiktoren im Modell fehlen, die bei der Vorhersage des Kriteriums eine Rolle spielen und nicht perfekt durch die enthaltenen Prädiktoren erklärt werden können
  - Unmöglich, sobald zufällige Messfehler vorliegen

Bias – Varianz Tradeoff

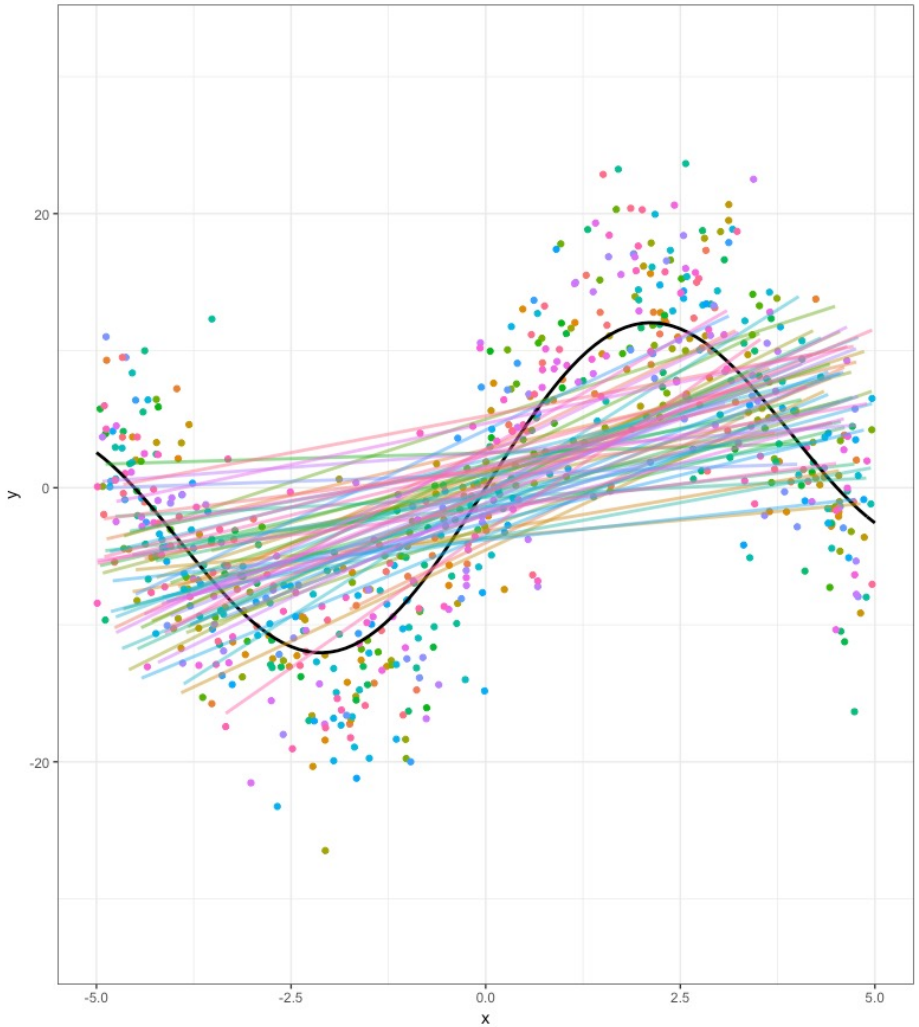
- Um einen niedrigen erwarteten Vorhersagefehler zu erzielen, muss die Kombination von Bias und Varianz möglichst niedrig sein. Dies bezeichnet man als Bias – Varianz Tradeoff.
- Unflexible Modellklassen haben tendenziell eine niedrige Varianz aber einen hohen Bias (z.B. multiple lineare Regression).
- Flexible Modellklassen haben tendenziell einen niedrigen Bias aber eine hohe Varianz (z.B. Entscheidungsbaum).
- Welcher Tradeoff für eine konkrete Anwendung optimal ist, hängt ab von der Art des wahren Zusammenhangs, der Höhe des nicht reduzierbaren Fehlers und der Größe der Stichprobe.
- Besonders leistungsstarke prädiktive Modellklassen erreichen sowohl einen niedrigen Bias, als auch eine niedrige Varianz (z.B. Random Forest).
  → Niedrigerer Vorhersagefehler in vielen Anwendungen

Bias – Varianz Tradeoff

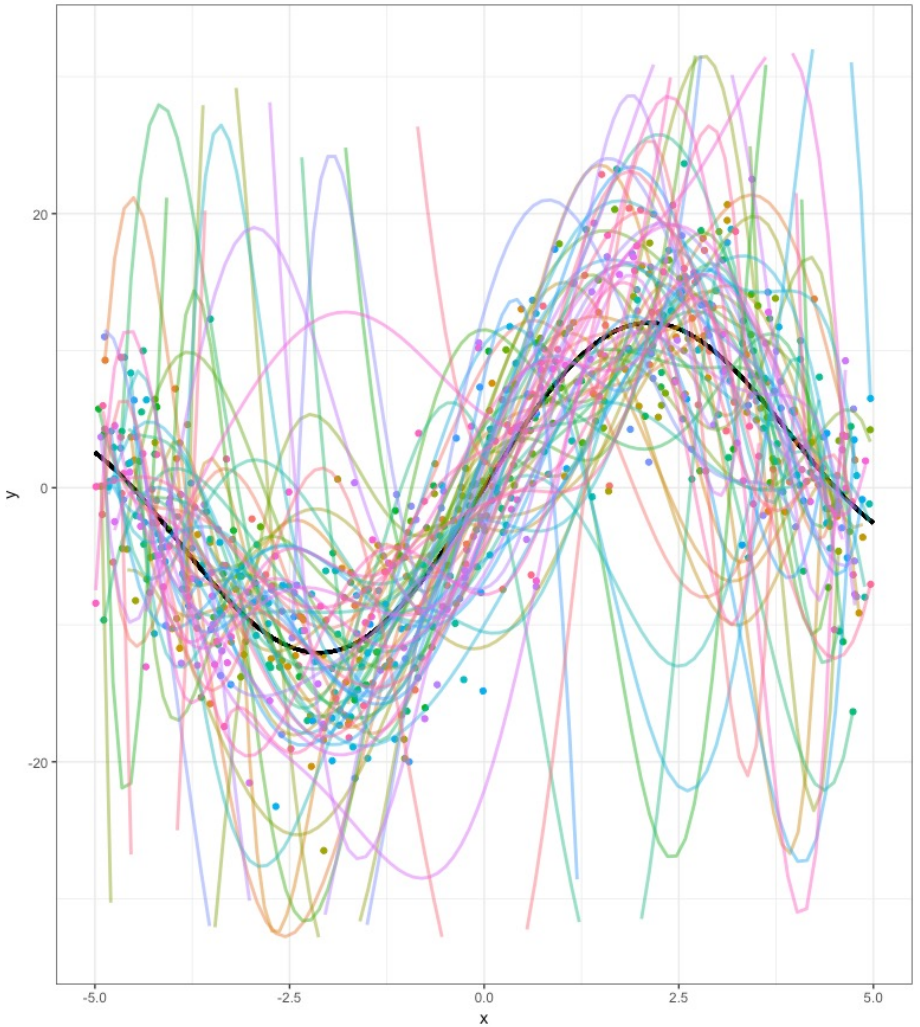


- 10 Stichproben der Größe N = 12 aus dem schwarzen Populationsmodell
- Trainiere für jede Stichprobe (Punkte der gleichen Farbe) ein unflexibles Modell (links), sowie ein flexibles Modell (rechts)

Bias – Varianz Tradeoff (Erhöhung der Anzahl der Stichproben)

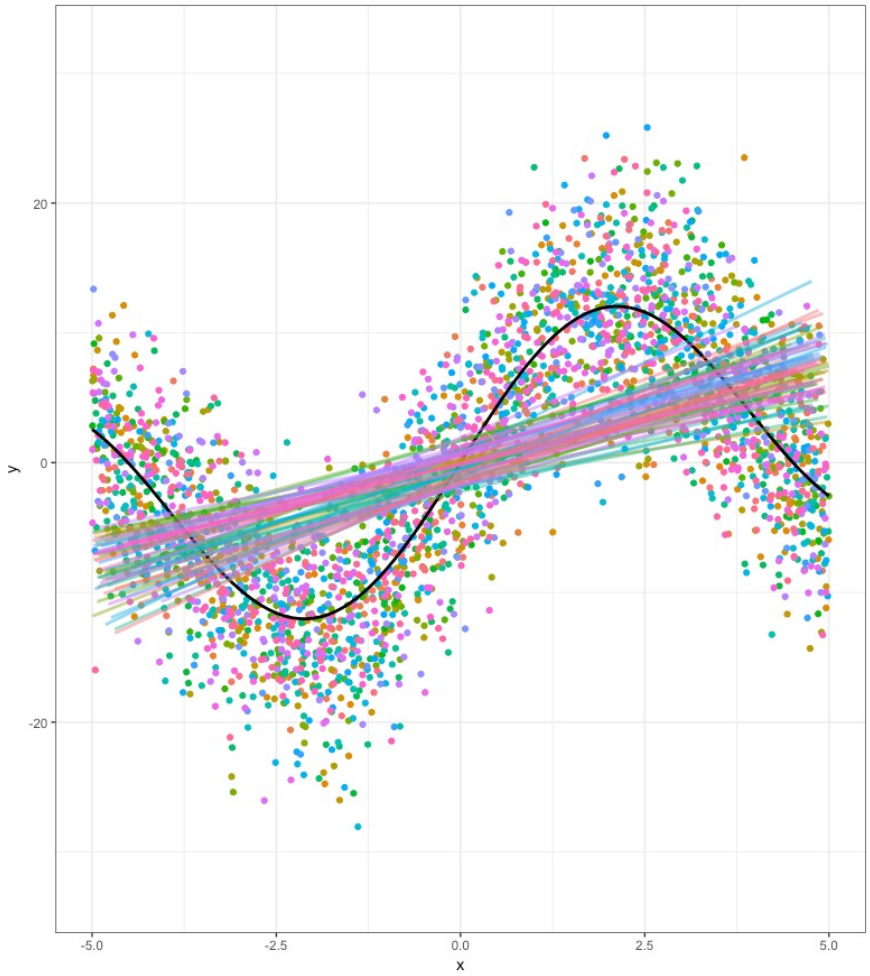


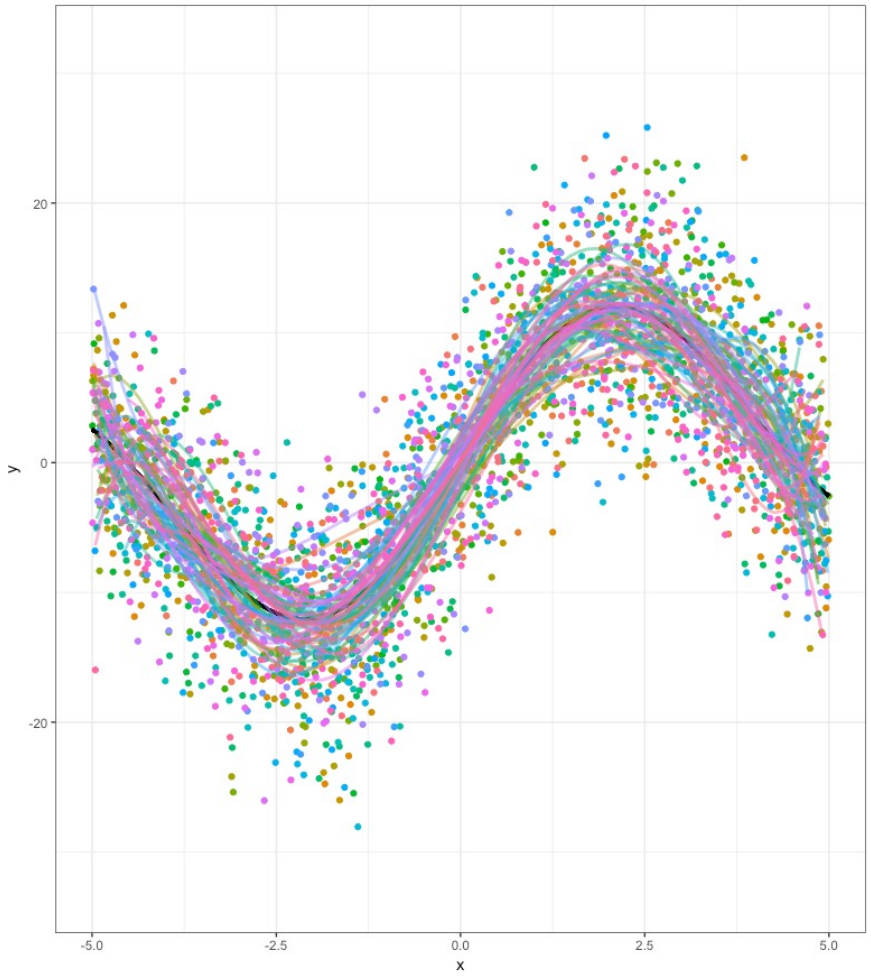
Bias: hoch
Varianz: niedrig



Bias: niedrig
Varianz: hoch

Bias – Varianz Tradeoff (Erhöhung der Stichprobengröße)





- Erhöhung der Stichprobengröße reduziert die Varianz (vor allem rechts)
- Damit deutlich genauere Vorhersagen mit dem flexibleren Modell

Agenda für heute

- Einführung zum prädiktiven Modell
  - Definition
  - Modellklassen
  - Annahmen
  - Regression vs. Klassifikation
  - Flexibilität & Interpretierbarkeit
- Quantifizierung der Vorhersagegüte
- Bias-Varianz Tradeoff

Nächste Woche

- Welche Rolle spielt Overfitting bei der Evaluation prädiktiver Modelle in der Praxis?
- Wie beurteilt man die Vorhersagegüte eines trainierten prädiktiven Modells anhand des gleichen Datensatzes, der auch für das Schätzen der Modellparameter verwendet wurde?
  → **Resampling Methoden**