# STAT 991: Topics in Modern Statistical Learning
# Lecture 1

Edgar Dobriban

January 18, 2022

# Overview

# Overview

## Course information

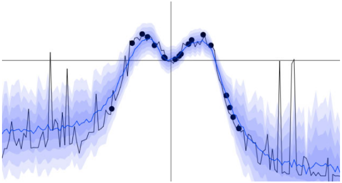
## Introduction to Topics

# Course information

- This is an advanced graduate class of seminar type

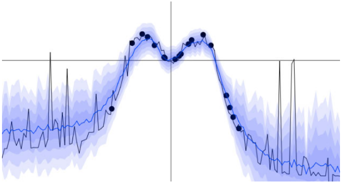
# Course information

- This is an advanced graduate class of seminar type
- The focus is on uncertainty quantification for machine learning methods: for instance conformal inference, prediction intervals/sets, calibration, OOD detection, etc.



# Course information

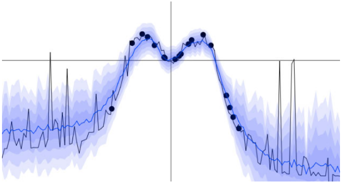
- This is an advanced graduate class of seminar type
- The focus is on uncertainty quantification for machine learning methods: for instance conformal inference, prediction intervals/sets, calibration, OOD detection, etc.



- All publicly available materials can be found on GitHub at https://github.com/dobriban/Topics-In-Modern-Statistical-Learning

# Course information

- This is an advanced graduate class of seminar type
- The focus is on uncertainty quantification for machine learning methods: for instance conformal inference, prediction intervals/sets, calibration, OOD detection, etc.



- All publicly available materials can be found on GitHub at `https://github.com/dobriban/Topics-In-Modern-Statistical-Learning`
- Some materials will be available only on Canvas (at least temporarily)

# Course information

- All lectures recorded. (At least) first two weeks online, rest currently planned in-person.

# Course information

- All lectures recorded. (At least) first two weeks online, rest currently planned in-person.
- Zoom lecture expectations

# Course information

- All lectures recorded. (At least) first two weeks online, rest currently planned in-person.
- Zoom lecture expectations
  - Mute yourself when not speaking

# Course information

- All lectures recorded. (At least) first two weeks online, rest currently planned in-person.
- Zoom lecture expectations
  - Mute yourself when not speaking
  - Ask questions as we go along: Unmute yourself and ask question (aim for interactivity)

# Course information

- All lectures recorded. (At least) first two weeks online, rest currently planned in-person.
- Zoom lecture expectations
  - Mute yourself when not speaking
  - Ask questions as we go along: Unmute yourself and ask question (aim for interactivity)
  - Camera on/off – up to you

# Course information

- Format:

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).
  - The goal of the course will also be to identity new research directions.

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).
  - The goal of the course will also be to identity new research directions.
  - Examples from past: three editions of Topics in Deep Learning: `https://github.com/dobriban/Topics-in-deep-learning`. Reseach topics came out: data augmentation, DeltaGrad.

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).
  - The goal of the course will also be to identity new research directions.
  - Examples from past: three editions of Topics in Deep Learning: `https://github.com/dobriban/Topics-in-deep-learning`. Reseach topics came out: data augmentation, DeltaGrad.
- As far I know, first course on this broad set of topics anywhere in the world. (sorry for the rough edges and omissions!)

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).
  - The goal of the course will also be to identity new research directions.
  - Examples from past: three editions of Topics in Deep Learning: `https://github.com/dobriban/Topics-in-deep-learning`. Reseach topics came out: data augmentation, DeltaGrad.
- As far I know, first course on this broad set of topics anywhere in the world. (sorry for the rough edges and omissions!)
- Goals:
  1. create an engaging, fun, but rigorous intellectual environment, where we will all learn and develop

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).
  - The goal of the course will also be to identity new research directions.
  - Examples from past: three editions of Topics in Deep Learning: `https://github.com/dobriban/Topics-in-deep-learning`. Reseach topics came out: data augmentation, DeltaGrad.
- As far I know, first course on this broad set of topics anywhere in the world. (sorry for the rough edges and omissions!)
- Goals:
  1. create an engaging, fun, but rigorous intellectual environment, where we will all learn and develop
  2. create broadly useful teaching materials

# Course information

- Format:
  - First several lectures by Edgar (and possible guests later)
  - Remaining lectures by students. We imagine a critical discussion of one or two papers per lecture; and several contiguous lectures on the same theme. The goal will be to develop a deep understanding of recent research (e.g., expect you to present proofs of results).
  - The goal of the course will also be to identity new research directions.
  - Examples from past: three editions of Topics in Deep Learning: `https://github.com/dobriban/Topics-in-deep-learning`. Reseach topics came out: data augmentation, DeltaGrad.
- As far I know, first course on this broad set of topics anywhere in the world. (sorry for the rough edges and omissions!)
- Goals:
  1. create an engaging, fun, but rigorous intellectual environment, where we will all learn and develop
  2. create broadly useful teaching materials
- See syllabus for: Prerequsites, Feedback, Grading Policy

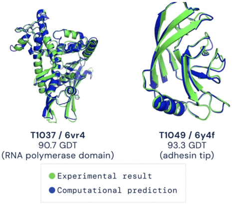
# Overview

Course information

Introduction to Topics

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")

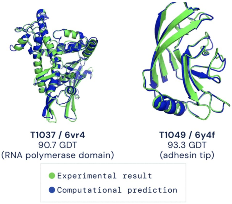



Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**

**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")



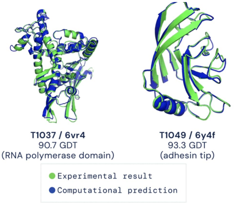

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**

**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

  - AlphaFold (Science's 2021 "breakthrough of the year")

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")



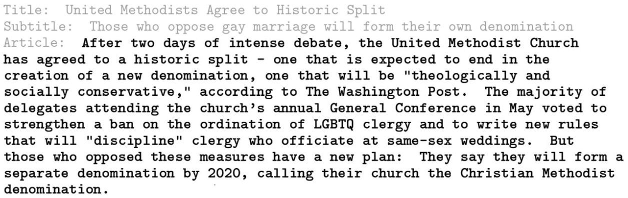
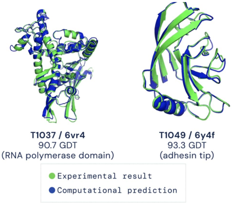


Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

  - AlphaFold (Science's 2021 "breakthrough of the year")
  - GPT-3

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")



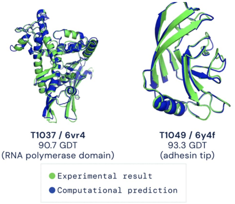

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**

**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

  - AlphaFold (Science's 2021 "breakthrough of the year")
  - GPT-3
  - Pro-level skin cancer classification

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")



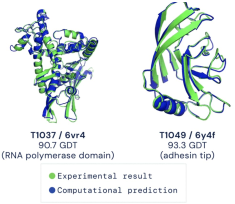

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**

**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

  - AlphaFold (Science's 2021 "breakthrough of the year")
  - GPT-3
  - Pro-level skin cancer classification
  - Self-driving cars

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")



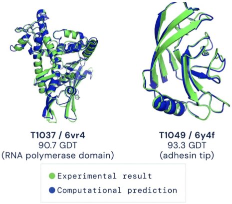

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**

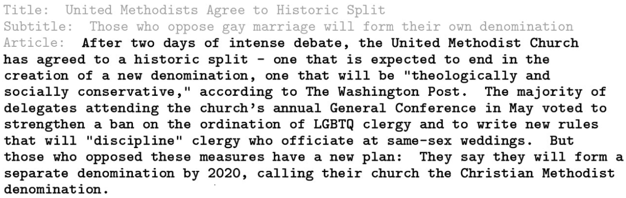
**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

  - AlphaFold (Science's 2021 "breakthrough of the year")
  - GPT-3
  - Pro-level skin cancer classification
  - Self-driving cars
  - Deep learning for stock prediction

# Perspective

- Machine learning (ML): one of the most salient intellectual developments in today's world (part of broader "Data Sciences")



Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

**Figure 3.14:** The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a written article (accuracy: 12%).

  - AlphaFold (Science's 2021 "breakthrough of the year")
  - GPT-3
  - Pro-level skin cancer classification
  - Self-driving cars
  - Deep learning for stock prediction
  - Causality and learning from natural experiments (a bit broader; Econ Nobel 2021)

# Perspective

- ML $=$ Prediction

# Perspective

- ML = Prediction
  - Given $x$ (inputs, features, covariates), predict $y$ (outputs, outcomes)

# Perspective

- ML = Prediction
  - Given $x$ (inputs, features, covariates), predict $y$ (outputs, outcomes)
- Examples

| | $x$ | $y$ |
|---|---|---|
| AlphaFold | Protein sequence | 3D structure |
| GPT-3 | Text prompt | Answer |
| Skin cancer | Skin image | Cancer type |
| Self-driving car | 3D environment, location, ... | Control signal |

# Perspective

- ML = Prediction
  - Given $x$ (inputs, features, covariates), predict $y$ (outputs, outcomes)
- Examples

| | $x$ | $y$ |
|---|---|---|
| AlphaFold | Protein sequence | 3D structure |
| GPT-3 | Text prompt | Answer |
| Skin cancer | Skin image | Cancer type |
| Self-driving car | 3D environment, location, ... | Control signal |

- Less glitzy, but very important examples: predicting phenotype from genetic variables, Predicting response to medical treatment

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:
  - GPT-3: given text prompt, ...?

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:
  - GPT-3: given text prompt, ...?
  - Self-driving cars: given 3D environment, location, ...?

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:
  - GPT-3: given text prompt, ...?
  - Self-driving cars: given 3D environment, location, ...?
  - AlphaFold: given protein sequence, ...?

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:
  - GPT-3: given text prompt, ...?
  - Self-driving cars: given 3D environment, location, ...?
  - AlphaFold: given protein sequence, ...?
  - Skin cancer classification: given skin image, ...?

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:
  - GPT-3: given text prompt, ...?
  - Self-driving cars: given 3D environment, location, ...?
  - AlphaFold: given protein sequence, ...?
  - Skin cancer classification: given skin image, ...?
- Standard ML pipeline (see next) does not provide a solution

# Uncertainty

- Given $x$, the answer $y$ is often not uniquely determined
  - even $x$ may be measured with error
- Examples of uncertainty:
  - GPT-3: given text prompt, ...?
  - Self-driving cars: given 3D environment, location, ...?
  - AlphaFold: given protein sequence, ...?
  - Skin cancer classification: given skin image, ...?
- Standard ML pipeline (see next) does not provide a solution
  - Currently no "standard" method

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]
- "Standard method": Empirical Risk Minimization (Wald, Vapnik, ...)

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]
- "Standard method": Empirical Risk Minimization (Wald, Vapnik, ...)
  - Class of predictors $\mathcal{H}$ (hypothesis class): linear models, deep nets, ...

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \dots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]
- "Standard method": Empirical Risk Minimization (Wald, Vapnik, ...)
  - Class of predictors $\mathcal{H}$ (hypothesis class): linear models, deep nets, ...
  - Loss function $\ell : \mathcal{Y} \times \mathcal{Y} \mapsto [0, \infty)$

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]
- "Standard method": Empirical Risk Minimization (Wald, Vapnik, ...)
  - Class of predictors $\mathcal{H}$ (hypothesis class): linear models, deep nets, ...
  - Loss function $\ell : \mathcal{Y} \times \mathcal{Y} \mapsto [0, \infty)$
  - Run some algorithm to approximately minimize empirical risk, finding $\hat{f}$

$$\min_{f \in H} R_n(f, Z) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i), y_i)$$

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]
- "Standard method": Empirical Risk Minimization (Wald, Vapnik, ...)
  - Class of predictors $\mathcal{H}$ (hypothesis class): linear models, deep nets, ...
  - Loss function $\ell : \mathcal{Y} \times \mathcal{Y} \mapsto [0, \infty)$
  - Run some algorithm to approximately minimize empirical risk, finding $\hat{f}$

$$\min_{f \in H} R_n(f, Z) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i), y_i)$$

  - Empirically evaluate: obtain new "test" dataset $Z' = \{(x_i', y_i'), i = 1, \ldots, n'\}$, find test loss $R_{n'}(\hat{f}, Z')$

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Standard ML pipeline

- For predicting $x \in \mathcal{X} \mapsto y \in \mathcal{Y}$,
  - collect a sample of "training" data $Z = \{(x_i, y_i),\ i = 1, \ldots, n\}$
  - find a predictor (hypothesis) $f : \mathcal{X} \mapsto \mathcal{Y}$ that "fits the data" well[1]
- "Standard method": Empirical Risk Minimization (Wald, Vapnik, ...)
  - Class of predictors $\mathcal{H}$ (hypothesis class): linear models, deep nets, ...
  - Loss function $\ell : \mathcal{Y} \times \mathcal{Y} \mapsto [0, \infty)$
  - Run some algorithm to approximately minimize empirical risk, finding $\hat{f}$

$$\min_{f \in H} R_n(f, Z) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i), y_i)$$

  - Empirically evaluate: obtain new "test" dataset $Z' = \{(x_i', y_i'), i = 1, \ldots, n'\}$, find test loss $R_{n'}(\hat{f}, Z')$
- Works... (perhaps surprisingly)

[1] Ideally, uncertainty quantification methods should not depend on fitting methods

# Uncertainty quantification

- Uncertainty = probability[2]

[2]what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions
- Examples:

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions
- Examples:
  - Upper Confidence Prediction: $P(y \leqslant f(x)) \geqslant 1 - \alpha$, for some $\alpha \in (0, 1)$

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions
- Examples:
  - Upper Confidence Prediction: $P(y \leqslant f(x)) \geqslant 1 - \alpha$, for some $\alpha \in (0, 1)$
  - Prediction Interval: $P(f_l(x) \leqslant y \leqslant f_u(x)) \geqslant 1 - \alpha$, for some $\alpha \in (0, 1)$

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions
- Examples:
  - Upper Confidence Prediction: $P(y \leqslant f(x)) \geqslant 1-\alpha$, for some $\alpha \in (0,1)$
  - Prediction Interval: $P(f_l(x) \leqslant y \leqslant f_u(x)) \geqslant 1-\alpha$, for some $\alpha \in (0,1)$
  - Prediction Set: for some mapping $C$ of inputs to subsets of $\mathcal{Y}$: $P(y \in C(x)) \geqslant 1-\alpha$, for some $\alpha \in (0,1)$. Or, conditional guarantee

$$P_Z(P_{(x,y)}(y \in C(x;Z)) \geqslant 1-\alpha) \geqslant 1-\delta$$

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions
- Examples:
  - Upper Confidence Prediction: $P(y \leqslant f(x)) \geqslant 1 - \alpha$, for some $\alpha \in (0, 1)$
  - Prediction Interval: $P(f_l(x) \leqslant y \leqslant f_u(x)) \geqslant 1 - \alpha$, for some $\alpha \in (0, 1)$
  - Prediction Set: for some mapping $C$ of inputs to subsets of $\mathcal{Y}$: $P(y \in C(x)) \geqslant 1 - \alpha$, for some $\alpha \in (0, 1)$. Or, conditional guarantee

$$P_Z(P_{(x,y)}(y \in C(x; Z)) \geqslant 1 - \alpha) \geqslant 1 - \delta$$

  - Calibration: for all appopriate $t$,

$$P(f(x) = y | p(x) = t) \approx t$$

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification

- Uncertainty = probability[2]
- Goal: Make probabilistic statements about predictions
- Examples:
  - Upper Confidence Prediction: $P(y \leqslant f(x)) \geqslant 1-\alpha$, for some $\alpha \in (0,1)$
  - Prediction Interval: $P(f_l(x) \leqslant y \leqslant f_u(x)) \geqslant 1-\alpha$, for some $\alpha \in (0,1)$
  - Prediction Set: for some mapping $C$ of inputs to subsets of $\mathcal{Y}$: $P(y \in C(x)) \geqslant 1-\alpha$, for some $\alpha \in (0,1)$. Or, conditional guarantee

$$P_Z(P_{(x,y)}(y \in C(x;Z)) \geqslant 1-\alpha) \geqslant 1-\delta$$

  - Calibration: for all appopriate $t$,

$$P(f(x) = y | p(x) = t) \approx t$$

  - Or... you name it: conditional on $x$, or on $y$, or only partial, or under dependence, or structured settings, or...
- Is uncertainty ever good?

[2] what is the meaning of "probability", or the source of randomness? discuss later

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?
- this is a deeper question than one might think. There is an entire area of the philosophy of probability and statistics.

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?
- this is a deeper question than one might think. There is an entire area of the philosophy of probability and statistics.
- For us the most common way of thinking is the frequency interpretation:

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?
- this is a deeper question than one might think. There is an entire area of the philosophy of probability and statistics.
- For us the most common way of thinking is the frequency interpretation:
  - $P(x \in A)$ is the long-term average over repeated experiments of the same type of the event that the observed quantity $x$ belongs to the set $A$

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?
- this is a deeper question than one might think. There is an entire area of the philosophy of probability and statistics.
- For us the most common way of thinking is the frequency interpretation:
  - $P(x \in A)$ is the long-term average over repeated experiments of the same type of the event that the observed quantity $x$ belongs to the set $A$
  - problems with this in ML?

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?
- this is a deeper question than one might think. There is an entire area of the philosophy of probability and statistics.
- For us the most common way of thinking is the frequency interpretation:
  - $P(x \in A)$ is the long-term average over repeated experiments of the same type of the event that the observed quantity $x$ belongs to the set $A$
  - problems with this in ML?
- Other interpretations: subjective beliefs (Dempster-Shafer), betting (Kelly), ...

# Uncertainty quantification: probability

- Probability: what does it mean? where does it come from?
- this is a deeper question than one might think. There is an entire area of the philosophy of probability and statistics.
- For us the most common way of thinking is the frequency interpretation:
  - $P(x \in A)$ is the long-term average over repeated experiments of the same type of the event that the observed quantity $x$ belongs to the set $A$
  - problems with this in ML?
- Other interpretations: subjective beliefs (Dempster-Shafer), betting (Kelly), ...
- The language of probability is already at the core of theoretical ML: generalization bounds, probabilistic ML (graphical models), Bayesian methods, etc. — so it is already widely accepted

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
    1. non-prediction/subjective problems, e.g., what is the best tasting food?

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?
- A1: So it makes most sense in some of the following settings

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?
- A1: So it makes most sense in some of the following settings
  1. objective prediction problems

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?
- A1: So it makes most sense in some of the following settings
  1. objective prediction problems
  2. there is no easily computed clear single answer

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?
- A1: So it makes most sense in some of the following settings
  1. objective prediction problems
  2. there is no easily computed clear single answer
  3. the stakes are somewhat high, and/or there is time to take into account the uncertainty – we care about this

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?
- A1: So it makes most sense in some of the following settings
  1. objective prediction problems
  2. there is no easily computed clear single answer
  3. the stakes are somewhat high, and/or there is time to take into account the uncertainty – we care about this
  4. there is data and computational resources

# Uncertainty quantification: why, when, how?

- Q1. Why/when do we need uncertainty quantification?
- There are problems where uncertainty quantification is almost moot:
  1. non-prediction/subjective problems, e.g., what is the best tasting food?
  2. deterministic problems, e.g., compute $3 + 4$
  3. low-stake problems where very quick decisions need to be taken, e.g., show ads on website
  4. problems with a lack of data or adequate computational resources, e.g., can I predict tomorrow's weather?
  5. very complex problems that depend on multiple domains, where prediction is inherently near-impossible, e.g., can NWS predict the weather 100 years in advance?
- A1: So it makes most sense in some of the following settings
  1. objective prediction problems
  2. there is no easily computed clear single answer
  3. the stakes are somewhat high, and/or there is time to take into account the uncertainty – we care about this
  4. there is data and computational resources
  5. relatively simple, isolated problem

# Uncertainty quantification: why, when, how?

- Q2. why via probability?

# Uncertainty quantification: why, when, how?

- Q2. why via probability?
- Sometimes one can/has to perform uncertainty quantification via other means

# Uncertainty quantification: why, when, how?

- Q2. why via probability?
- Sometimes one can/has to perform uncertainty quantification via other means
    1. Sometimes one can just provide a range, e.g., the high and low predicted precipitation for tomorrow, without any probabilistic interpretation

# Uncertainty quantification: why, when, how?

- Q2. why via probability?
- Sometimes one can/has to perform uncertainty quantification via other means
    1. Sometimes one can just provide a range, e.g., the high and low predicted precipitation for tomorrow, without any probabilistic interpretation
    2. Sometimes the data collection process is entirely out of the user's control, or it is not possible to formulate a probabilistic model, e.g., scraping web data

# Uncertainty quantification: why, when, how?

- Q2. why via probability?
- Sometimes one can/has to perform uncertainty quantification via other means
  1. Sometimes one can just provide a range, e.g., the high and low predicted precipitation for tomorrow, without any probabilistic interpretation
  2. Sometimes the data collection process is entirely out of the user's control, or it is not possible to formulate a probabilistic model, e.g., scraping web data
- A2: So some of the reasons/settings are

# Uncertainty quantification: why, when, how?

- Q2. why via probability?
- Sometimes one can/has to perform uncertainty quantification via other means
    1. Sometimes one can just provide a range, e.g., the high and low predicted precipitation for tomorrow, without any probabilistic interpretation
    2. Sometimes the data collection process is entirely out of the user's control, or it is not possible to formulate a probabilistic model, e.g., scraping web data
- A2: So some of the reasons/settings are
    1. If we want to be precise, and be able to use the answers for downstream calculations (e.g., to calculate an expected risk to a shipping company, we need probabilistic interpretation of the range of precipitation)

# Uncertainty quantification: why, when, how?

- Q2. why via probability?
- Sometimes one can/has to perform uncertainty quantification via other means
  1. Sometimes one can just provide a range, e.g., the high and low predicted precipitation for tomorrow, without any probabilistic interpretation
  2. Sometimes the data collection process is entirely out of the user's control, or it is not possible to formulate a probabilistic model, e.g., scraping web data
- A2: So some of the reasons/settings are
  1. If we want to be precise, and be able to use the answers for downstream calculations (e.g., to calculate an expected risk to a shipping company, we need probabilistic interpretation of the range of precipitation)
  2. In the sciences, a dominating paradigm is to collect a small amount of data in a controlled way, and make inferences based on that

# Uncertainty quantification: why, when, how?

- Q3. how via probability?

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)
  - Computer Science

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)
  - Computer Science
    - Conformal prediction (Vovk et al., '98-)

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)
  - Computer Science
    - Conformal prediction (Vovk et al., '98-)
    - The most popular current methods (Ensembles, Test-time model pertubation & augmentation, etc; Gal, Balasubramanian, ... )

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)
  - Computer Science
    - Conformal prediction (Vovk et al., '98-)
    - The most popular current methods (Ensembles, Test-time model pertubation & augmentation, etc; Gal, Balasubramanian, ... )
  - Applied areas

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)
  - Computer Science
    - Conformal prediction (Vovk et al., '98-)
    - The most popular current methods (Ensembles, Test-time model pertubation & augmentation, etc; Gal, Balasubramanian, ... )
  - Applied areas
    - Meteorology: calibration (aka validity, Miller, '63)

# Uncertainty quantification: why, when, how?

- Q3. how via probability?
- A3: Can start by following the most well established frameworks — There is well-developed scholarship, methods and results
  - Statistical inference and prediction
    - Confidence intervals (Neyman '37)
    - Prediction/tolerance regions (Wilks '41, Wald, Tukey 40's)
    - decision theoretic prediction (Takeuchi '75-'80)
  - Computer Science
    - Conformal prediction (Vovk et al., '98-)
    - The most popular current methods (Ensembles, Test-time model pertubation & augmentation, etc; Gal, Balasubramanian, ... )
  - Applied areas
    - Meteorology: calibration (aka validity, Miller, '63)
- There is no universally accepted approach. Some are mainly theoretically justified, some mainly empirically. Big open problem: can we have both?

# Uncertainty quantification: why, when, how?

- Q4: why now?

# Uncertainty quantification: why, when, how?

- Q4: why now?
- A4: Several reasons

# Uncertainty quantification: why, when, how?

- Q4: why now?
- A4: Several reasons
  - ML already successful in some cases, but still growing in popularity

# Uncertainty quantification: why, when, how?

- Q4: why now?
- A4: Several reasons
  - ML already successful in some cases, but still growing in popularity
  - there are some approaches to uncertainty quantification, but none fully satisfying.

# Uncertainty quantification: why, when, how?

- Q4: why now?
- A4: Several reasons
  - ML already successful in some cases, but still growing in popularity
  - there are some approaches to uncertainty quantification, but none fully satisfying.
  - One of the most exciting areas at the interface of statistics and ML. A lot of great work being done right now, by great research groups. You can be part of it!

# This course

- Go into any or all of the above in deep detail

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).
  - Roughly one lecture per student. Can work in teams of up to two, and then present two lectures together. Treat it as a course project.

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).
    - Roughly one lecture per student. Can work in teams of up to two, and then present two lectures together. Treat it as a course project.
    - Topics are suggested on the website (Mostly based on important or recent papers). Hopefully we can have some coherent structure (but do not aim for perfect).

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).
  - Roughly one lecture per student. Can work in teams of up to two, and then present two lectures together. Treat it as a course project.
  - Topics are suggested on the website (Mostly based on important or recent papers). Hopefully we can have some coherent structure (but do not aim for perfect).
  - There will be a Google sheet where you can sign up.

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).
  - Roughly one lecture per student. Can work in teams of up to two, and then present two lectures together. Treat it as a course project.
  - Topics are suggested on the website (Mostly based on important or recent papers). Hopefully we can have some coherent structure (but do not aim for perfect).
  - There will be a Google sheet where you can sign up.
  - Meet with Edgar one week in advance of your lecture to get feedback on your plan. Back and forth as you work on your presentation. Prepare cca 60-75 mins, to allow for discussion. Practice your presentation. Provide final presentation 24 hours in advance.

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).
  - Roughly one lecture per student. Can work in teams of up to two, and then present two lectures together. Treat it as a course project.
  - Topics are suggested on the website (Mostly based on important or recent papers). Hopefully we can have some coherent structure (but do not aim for perfect).
  - There will be a Google sheet where you can sign up.
  - Meet with Edgar one week in advance of your lecture to get feedback on your plan. Back and forth as you work on your presentation. Prepare cca 60-75 mins, to allow for discussion. Practice your presentation. Provide final presentation 24 hours in advance.
  - Can talk about your own work, But need to provide thorough introduction of the broader context, the technical tools, proofs etc. More of a lecture than a research talk.

# This course

- Go into any or all of the above in deep detail
- Overview by Edgar (mainly theory/methods), following lecture notes
- Followed by student lectures on topics of interest (can be theory/methods/algorithms/computation/applications...).
  - Roughly one lecture per student. Can work in teams of up to two, and then present two lectures together. Treat it as a course project.
  - Topics are suggested on the website (Mostly based on important or recent papers). Hopefully we can have some coherent structure (but do not aim for perfect).
  - There will be a Google sheet where you can sign up.
  - Meet with Edgar one week in advance of your lecture to get feedback on your plan. Back and forth as you work on your presentation. Prepare cca 60-75 mins, to allow for discussion. Practice your presentation. Provide final presentation 24 hours in advance.
  - Can talk about your own work, But need to provide thorough introduction of the broader context, the technical tools, proofs etc. More of a lecture than a research talk.
  - Aim to identify, discuss, and even answer, research questions.

# This course: to do now

- Website for topics: `https://github.com/dobriban/Topics-In-Modern-Statistical-Learning`

# This course: to do now

- Website for topics: `https://github.com/dobriban/Topics-In-Modern-Statistical-Learning`
- Google sheet to sign up for presentations: `https://docs.google.com/spreadsheets/d/1lYcUl1Z4pFdfQdJWzGQFmSkPqECblSnLuhBaugEJB10/edit?usp=sharing`
  - Sign up for one or multiple topics

# This course: to do now

- Website for topics: `https://github.com/dobriban/Topics-In-Modern-Statistical-Learning`
- Google sheet to sign up for presentations: `https://docs.google.com/spreadsheets/d/1lYcUl1Z4pFdfQdJWzGQFmSkPqECblSnLuhBaugEJB10/edit?usp=sharing`
  - Sign up for one or multiple topics
- Student introductions