Norm $\sim$ Dimensionality

Multiple Regimes in Learning

ICML - Haifa, Israel       June 24, 2010

Percy Liang         Nati Srebro
        UC Berkeley         TTI Chicago
A classic question
A classic question

Linear regression:

\[ X \sim \mathcal{N}(0, I_d) \quad Y = X^\top \beta^* + \mathcal{N}(0, 1) \]
A classic question

Linear regression:

\[ X \sim \mathcal{N}(0, I_d) \quad Y = X^\top \beta^* + \mathcal{N}(0, 1) \]

Norm: \( \|\beta^*\|_2 = C \)  
Dimensionality: \( d \)
A classic question

Linear regression:
\[ X \sim \mathcal{N}(0, I_d) \quad Y = X^\top \beta^* + \mathcal{N}(0, 1) \]

Norm: \( \|\beta^*\|_2 = C \)  
Dimensionality: \( d \)

What is the **excess risk**
\[ E_n = \mathbb{E}[(Y - X^\top \hat{\beta}_{\text{ERM on } n \text{ samples}})^2] - \mathbb{E}[(Y - X^\top \beta^*)^2]? \]
(equivalently, what is sample complexity \( n \) to achieve risk \( \epsilon \)?)
A classic question

Linear regression:
\[ X \sim \mathcal{N}(0, I_d) \quad Y = X^\top \beta^* + \mathcal{N}(0, 1) \]

Norm: \[ \|\beta^*\|_2 = C \]
Dimensionality: \[ d \]

What is the excess risk
\[ E_n = \mathbb{E}[(Y - X^\top \hat{\beta}_{\text{ERM on } n \text{ samples}})^2] - \mathbb{E}[(Y - X^\top \beta^*)^2] \]?
(equivalently, what is sample complexity \( n \) to achieve risk \( \epsilon \)?)

\[ E_n = O \left( \frac{C}{\sqrt{n}} \right) ? \]
A classic question

Linear regression:

\[ X \sim \mathcal{N}(0, I_d) \quad Y = X^\top \beta^* + \mathcal{N}(0, 1) \]

Norm: \[ \|\beta^*\|_2 = C \]

Dimensionality: \( d \)

What is the excess risk

\[ E_n = \mathbb{E}[(Y - X^\top \hat{\beta}_{\text{ERM on } n \text{ samples}})^2] - \mathbb{E}[(Y - X^\top \beta^*)^2] \]

(equivalently, what is sample complexity \( n \) to achieve risk \( \epsilon \)?)

\[ E_n = O \left( \frac{C}{\sqrt{n}} \right) ? \quad E_n = O \left( \frac{d}{n} \right) ? \]
Some answers

Finite sample complexity bounds (via uniform convergence)

\[ E_n = O \left( \frac{C}{\sqrt{n}} \right) \implies n = O(C^2) \]

independent of \( d \)

+ Works for any \( n \)  

- Bound can be loose
Some answers

Finite sample complexity bounds (via uniform convergence)

\[ E_n = O\left(\frac{C}{\sqrt{n}}\right) \Rightarrow n = O(C^2) \]

independent of \( d \)

+ Works for any \( n \)  

− Bound can be loose

Classical asymptotics (e.g., [van der Vaart, 1998; Liang & Jordan, 2008])

\[ E_n = \frac{d}{n} + o_p\left(\frac{1}{n}\right) \Rightarrow n = \Theta(d) \]

independent of \( C^2 \)

+ Exact up to first order  

− Works for large \( n \)
Some answers

Finite sample complexity bounds (via uniform convergence)

\[ E_n = O \left( \frac{C}{\sqrt{n}} \right) \quad \Rightarrow \quad n = O(C^2) \]

independent of \( d \)

+ Works for any \( n \)  

- Bound can be loose

Classical asymptotics (e.g., [van der Vaart, 1998; Liang & Jordan, 2008])

\[ E_n = \frac{d}{n} + o_p \left( \frac{1}{n} \right) \quad \Rightarrow \quad n = \Theta(d) \]

independent of \( C^2 \)

+ Exact up to first order  

- Works for large \( n \)

What’s the true behavior of \( E_n \)?
A heuristic answer

$E_n$

$\text{Trivial}$
A heuristic answer

\[ E_n \]

Trivial

Bounds

\[ n \]
A heuristic answer

$E_n$

Trivial

Bounds

Classical asymptotics

$n$
A heuristic answer

\[ E_n \]

Suggests multiple regimes
A heuristic answer

Suggests *multiple regimes*

Are these the actual regimes?

Where are the regime transitions and how do they behave?
The learning curve

Goal: precisely characterize the full learning curve at all $n$
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]

Asymptotics: \( n \to \infty \), exploit concentration to simplify
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]

Asymptotics: \( n \to \infty \), exploit concentration to simplify

Classical: \( \Psi \) fixed \( E_n \to 0 \)
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]

Asymptotics: \( n \to \infty \), exploit concentration to simplify

Classical: \( \Psi \text{ fixed } E_n \to 0 \)

High-dimensional: \( \Psi \to \infty \quad E_n > 0 \)
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]

Asymptotics: \( n \to \infty \), exploit concentration to simplify

Classical: \( \Psi \text{ fixed} \quad E_n \to 0 \)

High-dimensional: \( \Psi \to \infty \quad E_n > 0 \)

Important: preserve ratio between sample size and complexity

\( \Psi \to \tilde{\Psi} \quad (\text{e.g., } \frac{d}{n} \to \tilde{d}) \)
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]

Asymptotics: \( n \to \infty \), exploit concentration to simplify

Classical: \( \Psi \text{ fixed} \quad E_n \to 0 \)

High-dimensional: \( \Psi \to \infty \quad E_n > 0 \)

Important: preserve ratio between sample size and complexity

\[ \Psi \to \tilde{\Psi} \quad (e.g., \frac{d}{n} \to \tilde{d}) \]

Asymptotic excess risk:

\[ E_n(\Psi) \xrightarrow{P} \mathcal{E}(\tilde{\Psi}) \text{ non-degenerate} \]
Overview of approach

Excess risk:

\[ E_n(\Psi) \quad (\Psi \text{ is problem complexity, e.g., } \Psi = (C^2, d)) \]

Asymptotics: \( n \to \infty \), exploit concentration to simplify

Classical: \( \Psi \) fixed \( E_n \to 0 \)

High-dimensional: \( \Psi \to \infty \) \( E_n > 0 \)

Important: preserve ratio between sample size and complexity

\( \Psi \to \tilde{\Psi} \) (e.g., \( \frac{d}{n} \to \tilde{d} \))

Asymptotic excess risk:

\[ E_n(\Psi) \xrightarrow{P} \mathcal{E}(\tilde{\Psi}) \text{ non-degenerate} \]
Two examples

Mean estimation

Linear regression
Mean estimation: setup

Problem:

Data: $X^{(1)}, \ldots, X^{(n)} \sim \mathcal{N}(\mu^*, I)$

$$\|\mu^*\|^2 = B^2 \quad \mu^* \in \mathbb{R}^d$$

Goal: estimate $\mu^*$
Mean estimation: setup

Problem:

Data: \( X^{(1)}, \ldots, X^{(n)} \sim \mathcal{N}(\mu^*, I) \)

\[ \|\mu^*\|^2 = B^2 \quad \mu^* \in \mathbb{R}^d \]

Goal: estimate \( \mu^* \)

Estimator:

\[ \hat{\mu} = \frac{B \bar{X}}{\| \bar{X} \|}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X^{(i)} \]
Mean estimation: setup

Problem:

Data: \( X^{(1)}, \ldots, X^{(n)} \sim \mathcal{N}(\mu^*, I) \)

\[ \|\mu^*\|^2 = B^2 \quad \mu^* \in \mathbb{R}^d \]

Goal: estimate \( \mu^* \)

Estimator:

\[ \hat{\mu} = B \bar{X} / \|\bar{X}\|, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X^{(i)} \]

Excess risk:

\[ E_n = (\hat{\mu} - \mu^*)^2 \]
Mean estimation: theorem

If:

\[ B^2 \rightarrow \tilde{B}^2 \quad \frac{d}{n} \rightarrow \tilde{d} \]
Mean estimation: theorem

If:

\[ B^2 \to \tilde{B}^2 \quad \frac{d}{n} \to \tilde{d} \]

Then:

\[ E_n \xrightarrow{P} \mathcal{E}, \quad \mathcal{E} = 4\tilde{B}^2 \sin^2\left(\frac{1}{2} \arctan\sqrt{\frac{\tilde{d}}{\tilde{B}^2}}\right) \]
Mean estimation: theorem

If:

\[ B^2 \to \tilde{B}^2 \quad \frac{d}{n} \to \tilde{d} \]

Then:

\[ E_n \xrightarrow{P} \mathcal{E}, \quad \mathcal{E} = 4\tilde{B}^2 \sin^2 \left( \frac{1}{2} \arctan \sqrt{\frac{\tilde{d}}{\tilde{B}^2}} \right) \]

\[ B^2 = 1, \ d = 1000 \]
Mean estimation: theorem

If:

\[ B^2 \to \tilde{B}^2 \quad \frac{d}{n} \to \tilde{d} \]

Then:

\[ E_n \xrightarrow{P} \mathcal{E}, \quad \mathcal{E} = 4\tilde{B}^2 \sin^2 \left( \frac{1}{2} \arctan \sqrt{\frac{\tilde{d}}{\tilde{B}^2}} \right) \]

\[ B^2 = 1, \; d = 1000 \]
Mean estimation: theorem

If:

\[ B^2 \to \tilde{B}^2 \quad \frac{d}{n} \to \tilde{d} \]

Then:

\[ E_n \xrightarrow{P} \mathcal{E}, \quad \mathcal{E} = 4\tilde{B}^2 \sin^2 \left( \frac{1}{2} \arctan \sqrt{\frac{\tilde{d}}{\tilde{B}^2}} \right) \]

\[ \mathcal{E}^B = \min \left\{ 2\tilde{B}^2, \tilde{d} \right\} \]

\[ B^2 = 1, \quad d = 1000 \]
Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2$$ \hspace{1cm} \frac{d}{n} \to \tilde{d}

Then:

$$E_n \xrightarrow{P} \mathcal{E}, \quad \mathcal{E} = 4\tilde{B}^2 \sin^2 \left( \frac{1}{2} \arctan \sqrt{\frac{\tilde{d}}{\tilde{B}^2}} \right)$$

$$\mathcal{E}^B = \min \left\{ 2\tilde{B}^2, \tilde{d} \right\}$$

Random regime ($\tilde{B}^2 \ll \tilde{d}$):

$\hat{\mu}$ is a random guess on sphere

Norm $\tilde{B}$ dominates

---

$B^2 = 1, d = 1000$
Mean estimation: theorem

If:

\[ B^2 \to \tilde{B}^2 \quad \frac{d}{n} \to \tilde{d} \]

Then:

\[ E_n \xrightarrow{P} \mathcal{E}, \quad \mathcal{E} = 4\tilde{B}^2 \sin^2 \left( \frac{1}{2} \arctan \sqrt{\frac{\tilde{d}}{\tilde{B}^2}} \right) \]

\[ \mathcal{E}^B = \min \left\{ 2\tilde{B}^2, \tilde{d} \right\} \]

Random regime \((\tilde{B}^2 \ll \tilde{d})\):
\(\hat{\mu} \) is a random guess on sphere
Norm \(\tilde{B}\) dominates

Unregularized regime \((\tilde{d} \ll \tilde{B}^2)\):
\(\hat{\mu} - \mu^* \sim \text{Gaussian}\)
Dimensionality \(\tilde{d}\) dominates
Linear regression: setup

Data:

\[ X \sim \mathcal{N}(0, \Sigma) \quad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2) \]
Linear regression: setup

Data:
\[ X \sim \mathcal{N}(0, \Sigma) \quad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2) \]

Regularized least-squares estimator:
\[ \hat{\beta}^\lambda \overset{\text{def}}{=} \arg\min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \left( Y(i) - X(i)^\top \beta \right)^2 + \lambda \| \beta \|^2 \]
Linear regression: setup

Data:

\[ X \sim \mathcal{N}(0, \Sigma) \quad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2) \]

Regularized least-squares estimator:

\[ \hat{\beta}_\lambda \overset{\text{def}}{=} \arg\min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \left( Y^{(i)} - X^{(i)\top} \beta \right)^2 + \lambda \| \beta \|^2 \]

Excess risk (prediction squared loss):

\[ E_n^\lambda = \mathbb{E}[(Y - X^\top \hat{\beta}_\lambda)^2] - \mathbb{E}[(Y - X^\top \beta^*)^2] \]
Linear regression: setup

Data:

\[ X \sim \mathcal{N}(0, \Sigma) \quad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2) \]

Regularized least-squares estimator:

\[
\hat{\beta}^\lambda \overset{\text{def}}{=} \arg\min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \left( Y^{(i)} - X^{(i)^\top} \beta \right)^2 + \lambda \| \beta \|^2
\]

Excess risk (prediction squared loss):

\[
E_n^\lambda = \mathbb{E}[(Y - X^\top \hat{\beta}^\lambda)^2] - \mathbb{E}[(Y - X^\top \beta^*)^2]
\]

Oracle excess risk:

\[
E^*_n = \inf_{\lambda \geq 0} E_n^\lambda
\]
Linear regression: special structure

\[ \beta^* = (1, 0, \ldots, 0)^\top \]

\[ \Sigma = \text{diag}(B^2, \frac{C^2}{d-1}, \ldots, \frac{C^2}{d-1}) \]

Intuition:
- \( B \): norm of signal in data
- \( C \): norm of sum of irrelevant components
Linear regression: simplification

Componentwise estimator:

For $j = 1, \ldots, d$:

$$\hat{\beta}_j^\lambda = \text{least-squares solution using } \{X_j^{(i)}\}_{i=1}^n$$
Linear regression: simplification

Componentwise estimator:

For $j = 1, \ldots, d$:

$$\hat{\beta}_j^\lambda = \text{least-squares solution using } \{X_j^{(i)}\}_{i=1}^n$$

Note: oracle selection of $\lambda$ still couples components
Componentwise linear regression: theorem

If:

\[ B^2 \rightarrow \tilde{B}^2 \quad \frac{C^2 \sigma^2}{n} \rightarrow \tilde{C}'^2 \quad \frac{d \sigma^2}{n} \rightarrow \tilde{d} \]
Componentwise linear regression: theorem

If:

\[ B^2 \rightarrow \tilde{B}^2 \quad \frac{C^2 \sigma^2}{n} \rightarrow \tilde{C}^2 \quad \frac{d \sigma^2}{n} \rightarrow \tilde{d} \]

Then:

\[ E_n^* \xrightarrow{P} E^* \]
Componentwise linear regression: theorem

If:

\[ B^2 \rightarrow \tilde{B}^2 \quad \frac{C^2 \sigma^2}{n} \rightarrow \tilde{C}^2 \quad \frac{d\sigma^2}{n} \rightarrow \tilde{d} \]

Then:

\[ E^*_n = \inf_{\lambda \geq 0} E^\lambda_n \overset{P}{\rightarrow} E^* \]
Componentwise linear regression: theorem

If:

\[ B^2 \rightarrow \tilde{B}^2 \quad \frac{c^2 \sigma^2}{n} \rightarrow \tilde{C}^2 \quad \frac{d\sigma^2}{n} \rightarrow \tilde{d} \]

Then:

\[ E_n^* = \inf_{\lambda \geq 0} E_n^\lambda \xrightarrow{P} \inf_{\lambda \geq 0} E^\lambda = E^* \]
Componentwise linear regression: theorem

If:

\[ B^2 \to \tilde{B}^2 \quad \frac{C^2 \sigma^2}{n} \to \tilde{C}^2 \quad \frac{d \sigma^2}{n} \to \tilde{d} \]

Then:

\[ E_n^* = \inf_{\lambda \geq 0} E_n^\lambda \xrightarrow{P} \inf_{\lambda \geq 0} E^\lambda = E^* \]

\[ E^\lambda = \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4 \tilde{d}}{(\tilde{C}^2 \tilde{d} + \lambda)^2} \]

\[ \text{squared bias} \quad \text{variance} \]
Componentwise linear regression: intuition

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4}{\tilde{d}} \left( \frac{\tilde{C}^2}{\tilde{d}} + \lambda \right)^2$$

- Squared bias
- Variance
Componentwise linear regression: intuition

\[ \mathcal{E}^* = \inf_{\lambda \geq 0} \left( \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} \right) + \frac{\tilde{C}^4}{\tilde{d}} \left( \frac{\tilde{C}^2}{\tilde{d}} + \lambda \right)^2 \]

Squared bias

Variance

\[ B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100 \]
Componentwise linear regression: intuition

\[ E^* = \inf_{\lambda \geq 0} \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4}{\tilde{d}^2} \]

- squared bias
- variance

**Diagram:**
- Excess risk vs. sample size
- Actual (\( E_n \))
- Asymptotic (\( E^* \))

**Equation Parameters:**
- \( B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100 \)
Componentwise linear regression: intuition

\[
\mathcal{E}^* = \inf_{\lambda \geq 0} \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4}{\tilde{d}}
\]

\[
\text{squared bias} \quad \text{variance}
\]

\[
\mathcal{E}_B \overset{\text{def}}{=} \min \left\{ \tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B}\sqrt{\tilde{d}}}, \tilde{d} \right\}
\]

For the given values:

\[
B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100
\]
Componentwise linear regression: intuition

\[ \mathcal{E}^* = \inf_{\lambda \geq 0} \left( \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4}{\tilde{d}} \right) \]

Squared bias + Variance

\[ \mathcal{E}_B \overset{\text{def}}{=} \min \left\{ \tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B} \sqrt{\tilde{d}}}, \tilde{d} \right\} \]

Random (\( \lambda \to \infty \)): Bias dominates; \( \mathcal{E}^* \sim 1 \)

\[ B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100 \]
Componentwise linear regression: intuition

\[ \mathcal{E}^* = \inf_{\lambda \geq 0} \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4}{\tilde{d}} \]

squared bias

variance

\[ \mathcal{E}^B \overset{\text{def}}{=} \min \left\{ \frac{\tilde{B}^2}{\tilde{B} \sqrt{\tilde{d}}}, \frac{2\tilde{C}^2}{\tilde{d}} \right\} \]

Random (\( \lambda \to \infty \)):
Bias dominates; \( \mathcal{E}^* \sim 1 \)

Regularized (\( \lambda \) non-trivial):
Balance bias/variance; \( \mathcal{E}^* \sim \frac{1}{\sqrt{n}} \)

\( B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100 \)
Componentwise linear regression: intuition

\[ \mathcal{E}^* = \inf_{\lambda \geq 0} \frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2} + \frac{\tilde{C}^4}{\tilde{d}} \]

- squared bias
- variance

\[ \mathcal{E}^B \overset{\text{def}}{=} \min \left\{ \tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B}\sqrt{\tilde{d}}} \right\} \]

**Random (\( \lambda \to \infty \))**: Bias dominates; \( \mathcal{E}^* \sim 1 \)

**Regularized (\( \lambda \) non-trivial)**: Balance bias/variance; \( \mathcal{E}^* \sim \frac{1}{\sqrt{n}} \)

**Unregularized (\( \lambda \to 0 \))**: Variance dominates; \( \mathcal{E}^* \sim \frac{1}{n} \)

\[ B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100 \]
Full linear regression: speculation

Stitch together results:

$$E^* \approx \min \{ \tilde{B}^2, \text{trivial} \}$$
Full linear regression: speculation

Stitch together results:

\[ E^* \approx \min \left\{ \tilde{B}^2, \quad O \left( \frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C} \right) \right\}, \]

trivial bounds from [Srebro et al., 2010]
Full linear regression: speculation

Stitch together results:

$$\mathcal{E}^* \approx \min\{ \tilde{B}^2, \ O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right), \ \tilde{d} \}$$

trivial bounds from \[Srebro et al., 2010\] classical asymptotics
Full linear regression: speculation

Stitch together results:

\[ \mathcal{E}^* \approx \min \left\{ \tilde{B}^2, \quad O \left( \frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C} \right), \quad \tilde{d} \right\} \]

trivial bounds from [Srebro et al., 2010] classical asymptotics

Four regimes:
Full linear regression: speculation

Stitch together results:

\[ E^* \approx \min \{ \tilde{B}^2, O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right), \tilde{d} \} \]

trivial bounds from [Srebro et al., 2010]
classical asymptotics

Four regimes:

- **Random**: \[ E_n \approx B^2 \]
Full linear regression: speculation

Stitch together results:

\[ E^* \approx \min \{ \tilde{B}^2, O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right), \tilde{d} \} \]

trivial bounds from [Srebro et al., 2010]
classical asymptotics

Four regimes:

Random \( E_n \approx B^2 \)

Low noise \( E^* \approx C^2_n \)

\[ \frac{C^2}{B^2} \]

\[ \frac{C^2}{\sigma^2} \]
Full linear regression: speculation

Stitch together results:

\[ E^* \approx \min \{ \tilde{B}^2, O \left( \frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C} \right), \tilde{d} \} \]

trivial bounds from [Srebro et al., 2010] classical asymptotics

Four regimes:

- Random: \( E^* \approx B^2 \)
- Low noise: \( E^* \approx \frac{C^2}{n} \)
- Regularized: \( E^* \approx \sqrt{\frac{C^2\sigma^2}{n}} \)
Full linear regression: speculation

Stitch together results:

$$\mathcal{E}^* \approx \min \{ \tilde{B}^2, \ O \left( \tilde{C}^2 \tilde{\sigma}^2 + \tilde{C} \right), \ \tilde{d} \}$$

trivial

bounds from [Srebro et al., 2010]

classical

asymptotics

Four regimes:

- **Random**
  $$\mathcal{E}^* \approx B^2$$

- **Low noise**
  $$\mathcal{E}^* \approx \frac{C^2}{n}$$

- **Regularized**
  $$\mathcal{E}^* \approx \sqrt{\frac{C^2\sigma^2}{n}}$$

- **Unregularized**
  $$\mathcal{E}^* \approx \frac{d\sigma^2}{n}$$
Full linear regression: speculation

Stitch together results:

\[ E^* \approx \min\{ \tilde{B}^2, O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right), \tilde{d} \} \]

trivial bounds from [Srebro et al., 2010] classical asymptotics

Four regimes:

Random: \[ E^* \approx B^2 \]
Low noise: \[ E^* \approx \frac{C^2}{\sigma^2} \]
Regularized: \[ E^* \approx \sqrt{\frac{C^2\sigma^2}{n}} \]
Unregularized: \[ E^* \approx \frac{d\sigma^2}{n} \]
Conclusions

Summary: studied two simple examples
(mean estimation and componentwise linear regression)
Conclusions

**Summary:** studied two simple examples
(mean estimation and componentwise linear regression)

**Observations:**
Complexities (norm, dimensionality) active in different regimes
Transitions between regimes are smooth
Conclusions

Summary: studied two simple examples
(mean estimation and componentwise linear regression)

Observations:
Complexities (norm, dimensionality) active in different regimes
Transitions between regimes are smooth

Broad goal: understand the learning curve
Conclusions

Summary: studied two simple examples
(mean estimation and componentwise linear regression)

Observations:
Complexities (norm, dimensionality) active in different regimes
Transitions between regimes are smooth

Broad goal: understand the learning curve
Finite sample bounds, classical asymptotics: partial picture
High-dimensional asymptotics (statistics, statistical physics)
Key: concentration in high dimensions
Conclusions

Summary: studied two simple examples
(mean estimation and componentwise linear regression)

Observations:
Complexities (norm, dimensionality) active in different regimes
Transitions between regimes are smooth

Broad goal: understand the learning curve
Finite sample bounds, classical asymptotics: partial picture
High-dimensional asymptotics (statistics, statistical physics)
Key: concentration in high dimensions

Future work: analyze more complex settings