Norm $\sim$ Dimensionality

Multiple Regimes in Learning

ICML - Haifa, Israel June 24, 2010

Percy Liang
UC Berkeley

Nati Srebro
TTI Chicago

A classic question

A classic question

Linear regression:

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

A classic question

Linear regression:

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

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

A classic question

Linear regression:

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

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 ϵ?)

A classic question

Linear regression:

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

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 ϵ?)

$$E_n = O\left(\frac{C}{\sqrt{n}}\right)?$$

A classic question

Linear regression:

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

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 ϵ?)

$$E_n = O\left(\frac{C}{\sqrt{n}}\right)? \qquad E_n = O\left(\frac{d}{n}\right)?$$

Some answers

Finite sample complexity bounds (via uniform convergence)

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

independent of d

$+$ Works for any n $\qquad$ $-$ Bound can be loose

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 $\qquad$ $-$ 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 $\qquad$ $-$ 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 $\qquad$ $-$ Bound can be loose

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

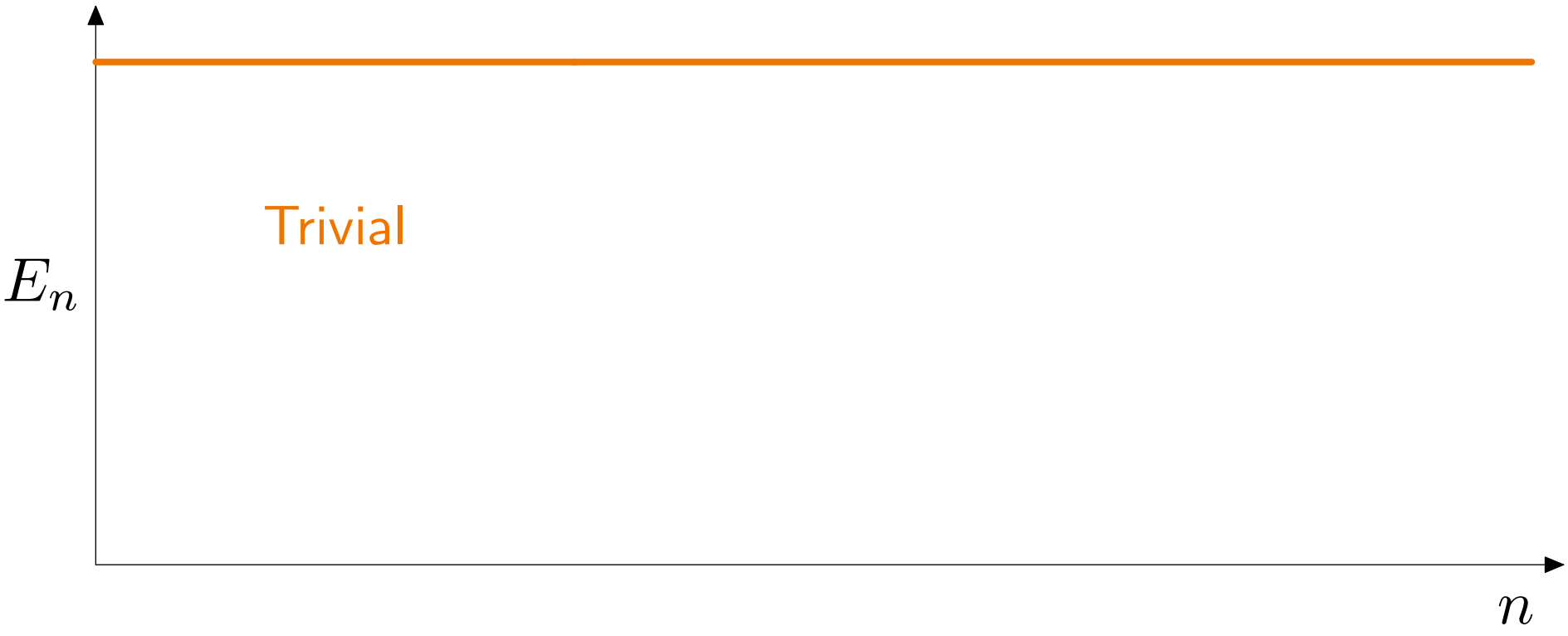
$$E_n = \frac{d}{n} + o_p\left(\frac{1}{n}\right) \quad \Rightarrow \quad n = \Theta(d)$$

independent of C^2

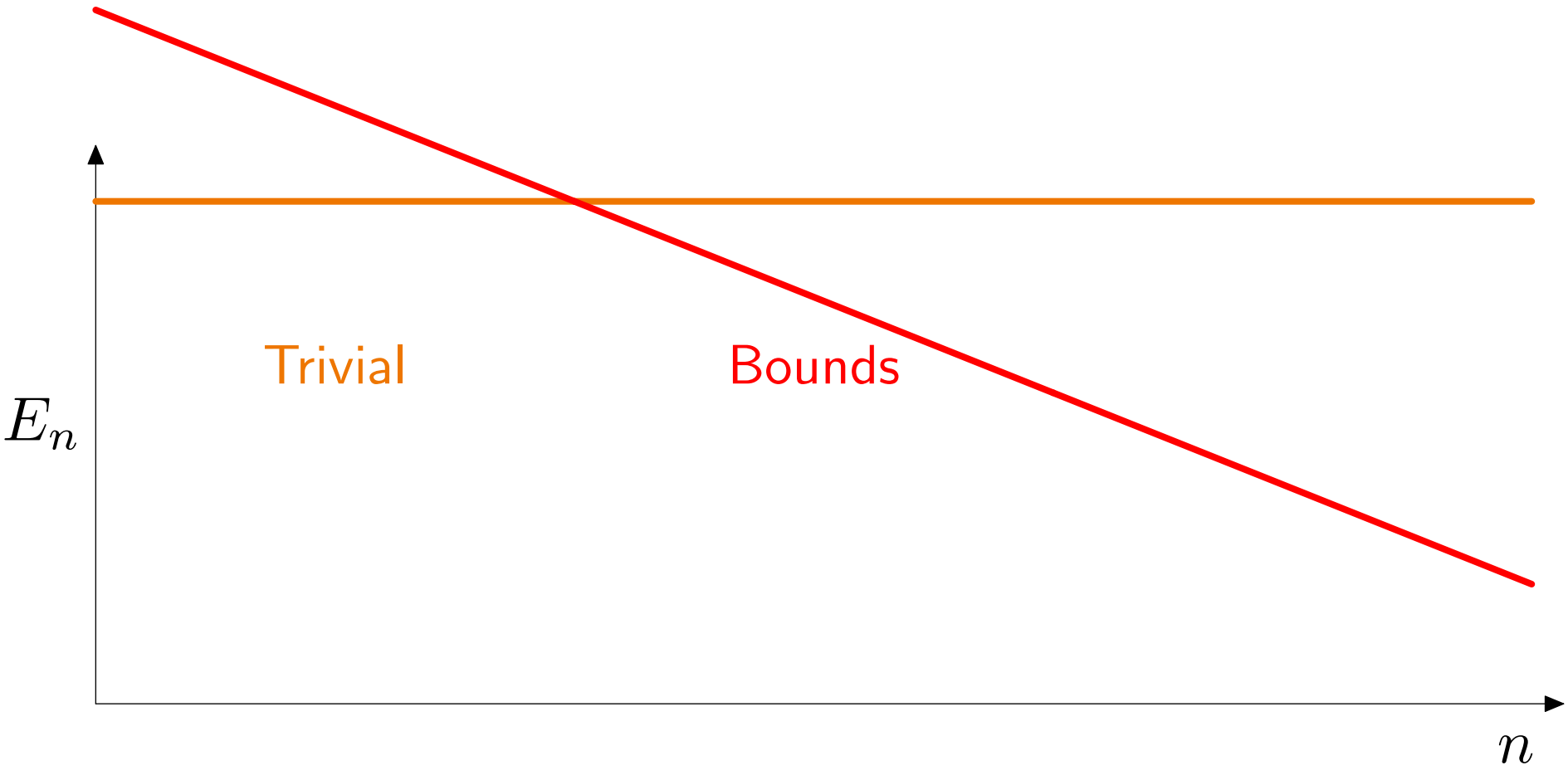
$+$ Exact up to first order $\qquad$ $-$ Works for large n

What's the true behavior of E_n?

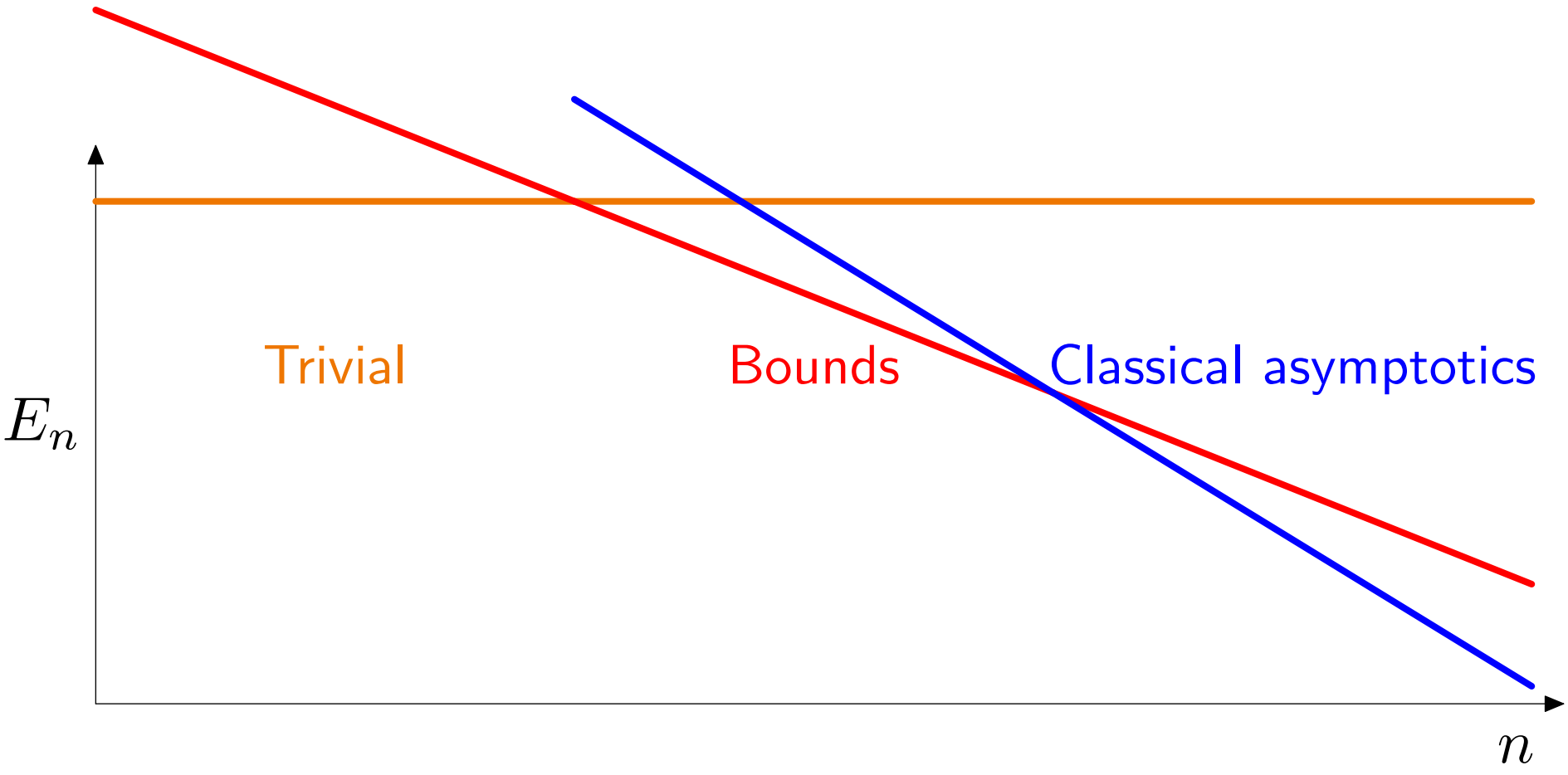
A heuristic answer



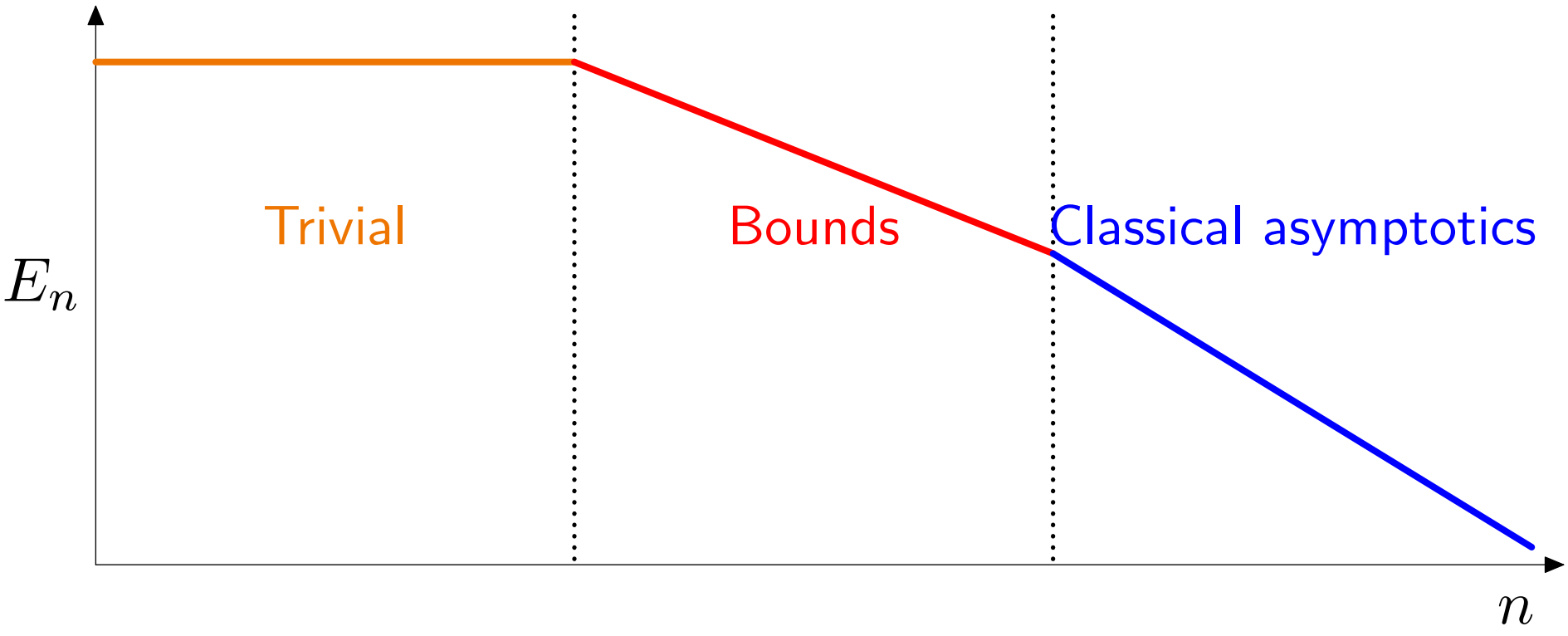
A heuristic answer



A heuristic answer

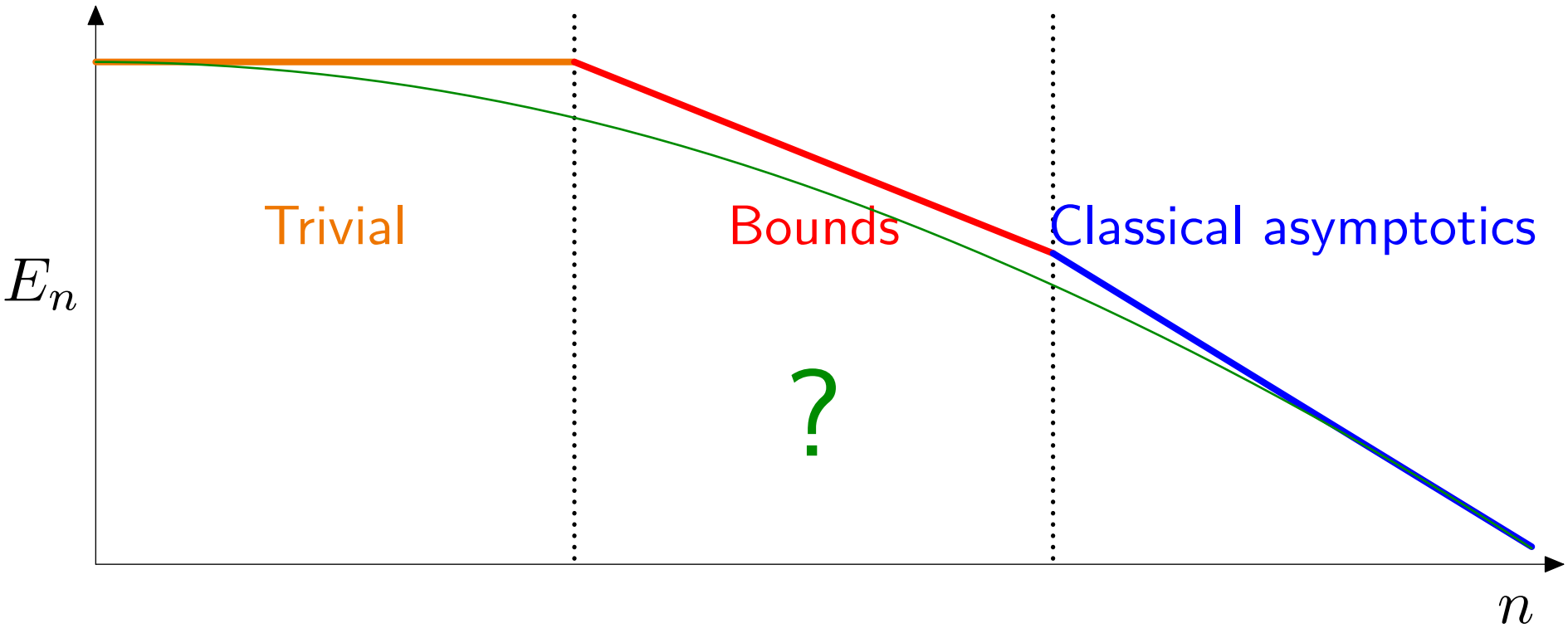


A heuristic answer



Suggests **multiple regimes**

A heuristic answer



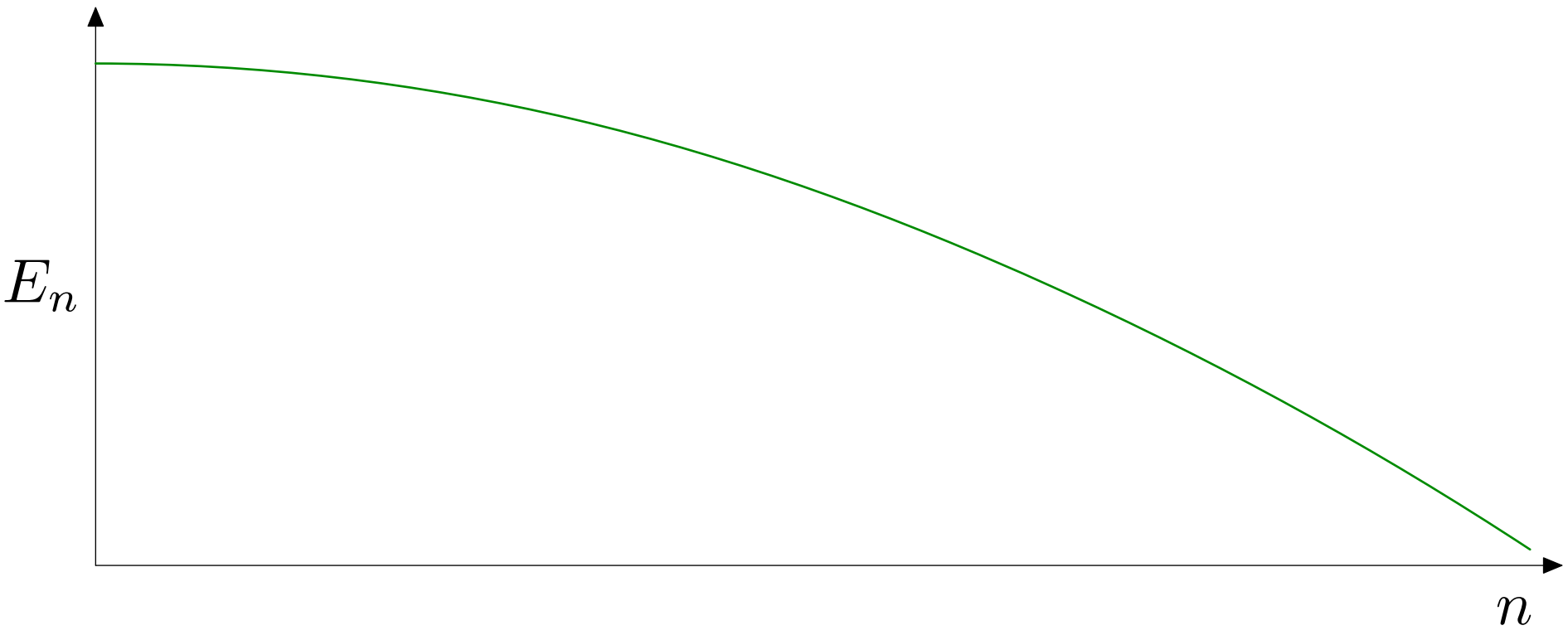
Suggests **multiple regimes**

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: Ψ fixed $\quad E_n \to 0$

Overview of approach

Excess risk:

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

Asymptotics: $n \to \infty$, exploit concentration to simplify

Classical:	Ψ fixed	$E_n \to 0$
High-dimensional:	$\Psi \to \infty$	$E_n > 0$

Overview of approach

Excess risk:

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

Asymptotics: $n \to \infty$, exploit concentration to simplify

Classical:	Ψ fixed	$E_n \to 0$
High-dimensional:	$\Psi \to \infty$	$E_n > 0$

Important: preserve ratio between sample size and complexity

$$\Psi \to \tilde{\Psi} \qquad (\text{e.g., } \tfrac{d}{n} \to \tilde{d})$$

Overview of approach

Excess risk:

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

Asymptotics: $n \to \infty$, exploit concentration to simplify

Classical:	Ψ fixed	$E_n \to 0$
High-dimensional:	$\Psi \to \infty$	$E_n > 0$

Important: preserve ratio between sample size and complexity

$$\Psi \to \tilde{\Psi} \qquad (\text{e.g., } \tfrac{d}{n} \to \tilde{d})$$

Asymptotic excess risk:

$$E_n(\Psi) \xrightarrow{P} \mathcal{E}(\tilde{\Psi}) \text{ non-degenerate}$$

Overview of approach

Excess risk:

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

Asymptotics: $n \to \infty$, exploit concentration to simplify

Classical:	Ψ fixed	$E_n \to 0$
High-dimensional:	$\Psi \to \infty$	$E_n > 0$

Important: preserve ratio between sample size and complexity

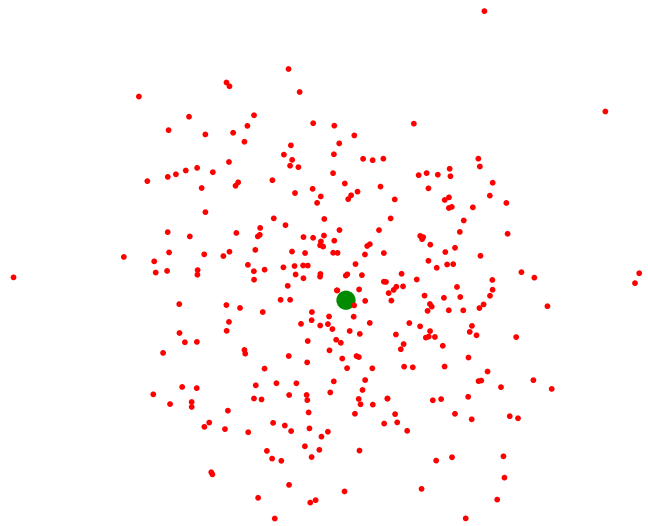
$$\Psi \to \tilde{\Psi} \qquad (\text{e.g., } \tfrac{d}{n} \to \tilde{d})$$

Asymptotic excess risk:

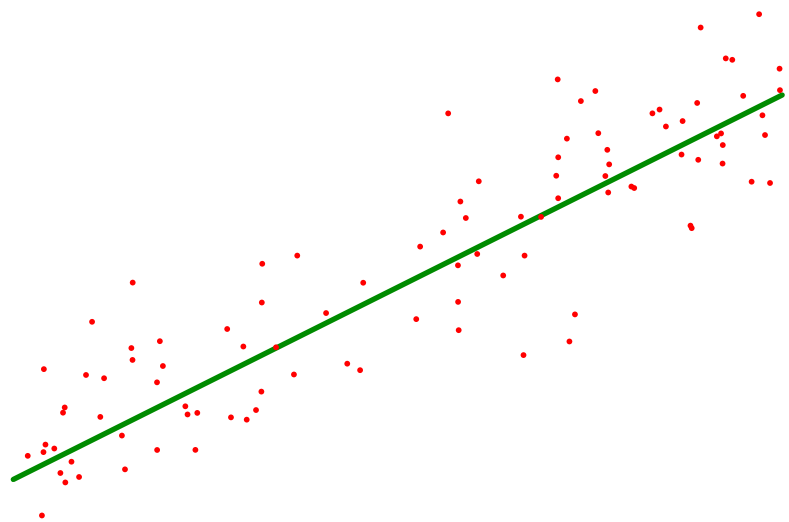
$$E_n(\Psi) \xrightarrow{P} \boxed{\mathcal{E}(\tilde{\Psi})} \text{ non-degenerate}$$

Two examples

Mean estimation



Linear regression



Mean estimation: setup

Problem:

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

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

Goal: estimate μ^*

Mean estimation: setup

Problem:

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

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

Goal: estimate μ^*

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

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

Goal: estimate μ^*

Estimator:

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

Excess risk:

$$E_n = (\hat{\mu} - \mu^*)^2$$

Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Then:

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

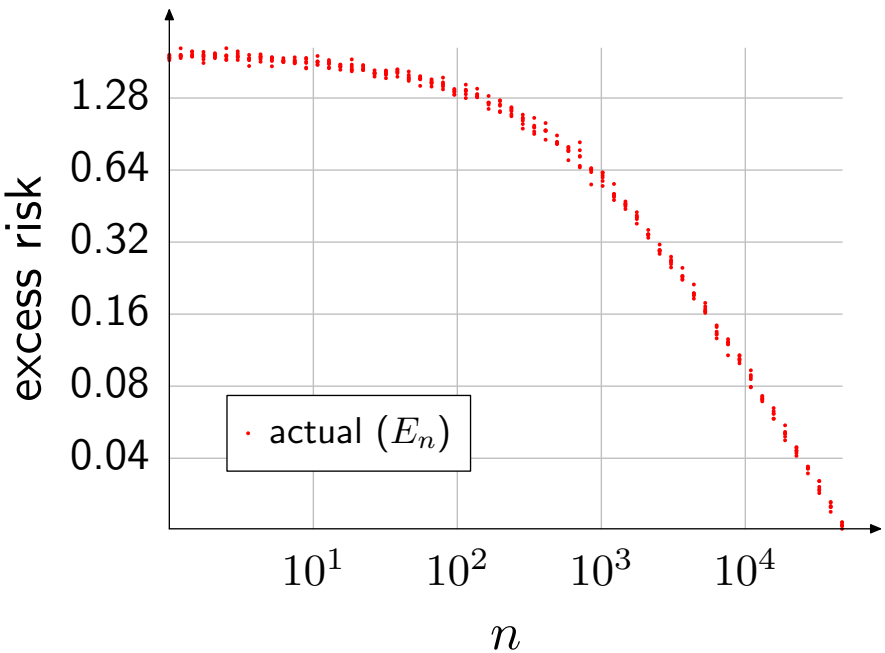
Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Then:

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



$B^2 = 1, d = 1000$

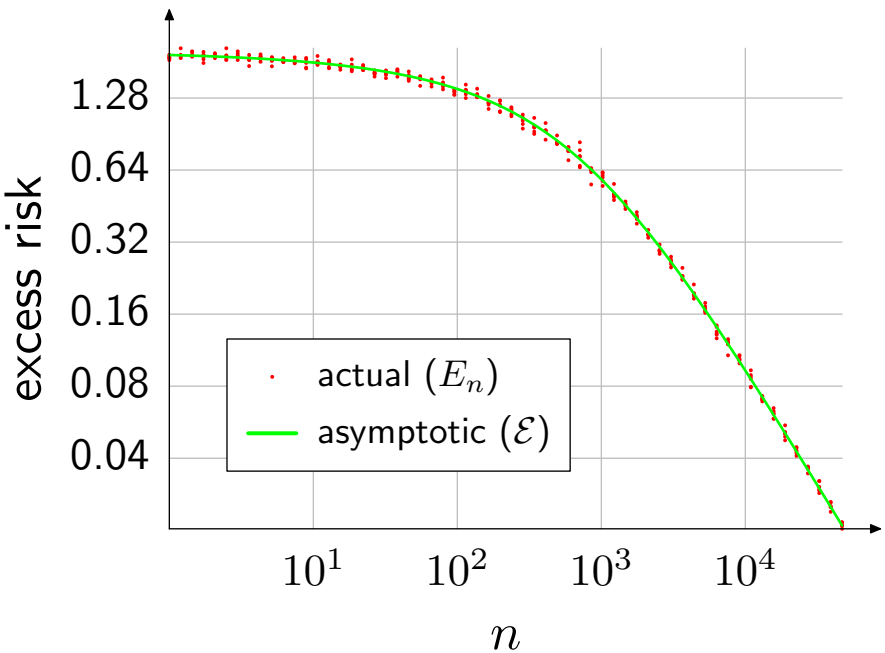
Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Then:

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



$B^2 = 1, d = 1000$

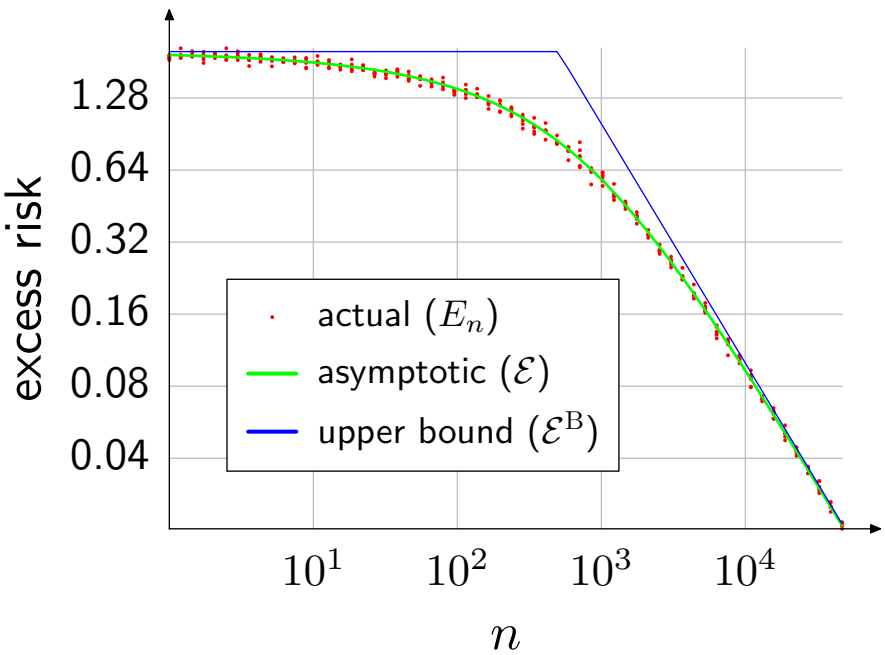
Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Then:

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



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

$B^2 = 1, d = 1000$

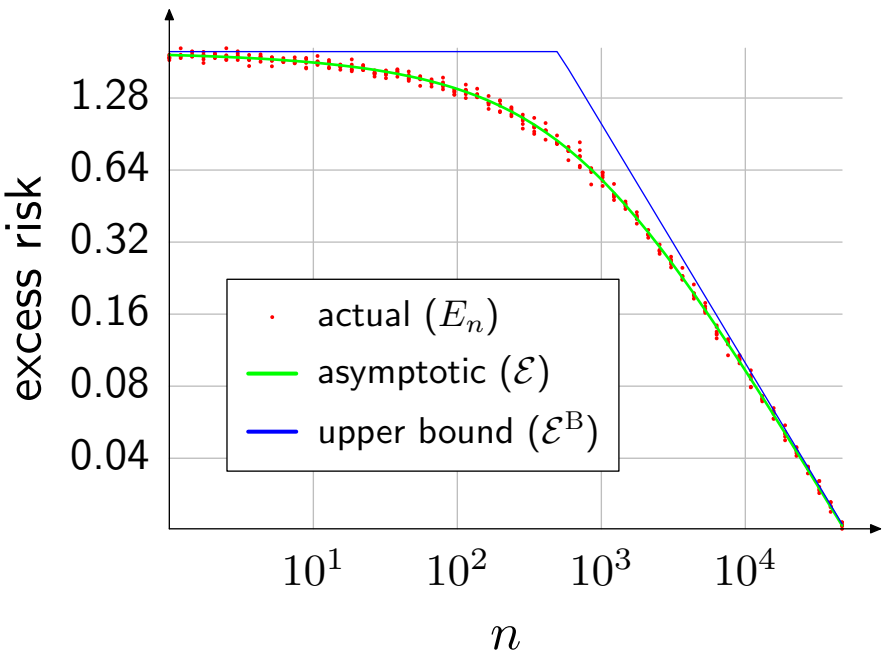
Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Then:

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



$B^2 = 1, d = 1000$

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

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

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

Norm $\tilde{B}$ dominates

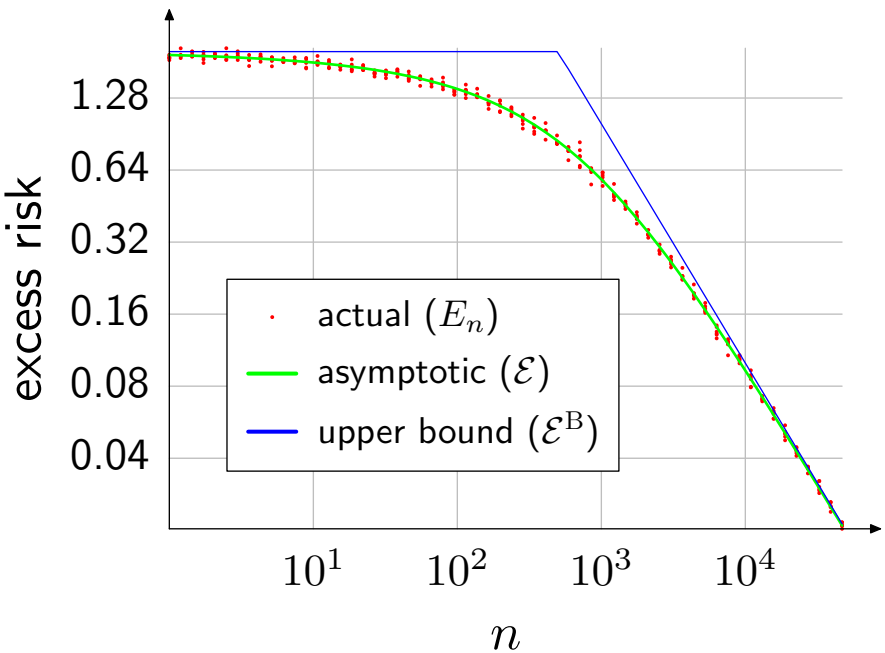
Mean estimation: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{d}{n} \to \tilde{d}$$

Then:

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



$B^2 = 1, d = 1000$

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

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

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

Norm $\tilde{B}$ dominates

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

$\hat{\mu} - \mu^* \sim$ Gaussian

Dimensionality $\tilde{d}$ dominates

Linear regression: setup

Data:

$$X \sim \mathcal{N}(0, \Sigma) \qquad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2)$$

Linear regression: setup

Data:

$$X \sim \mathcal{N}(0, \Sigma) \qquad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2)$$

Regularized least-squares estimator:

$$\hat{\beta}^\lambda \stackrel{\text{def}}{=} \underset{\beta \in \mathbb{R}^d}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(Y^{(i)} - X^{(i)\top} \beta \right)^2 + \lambda \|\beta\|^2$$

Linear regression: setup

Data:

$$X \sim \mathcal{N}(0, \Sigma) \qquad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2)$$

Regularized least-squares estimator:

$$\hat{\beta}^\lambda \stackrel{\text{def}}{=} \underset{\beta \in \mathbb{R}^d}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(Y^{(i)} - X^{(i)\top} \beta \right)^2 + \lambda \|\beta\|^2$$

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:

$$X \sim \mathcal{N}(0, \Sigma) \qquad Y \sim \mathcal{N}(X^\top \beta^*, \sigma^2)$$

Regularized least-squares estimator:

$$\hat{\beta}^\lambda \stackrel{\text{def}}{=} \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \left(Y^{(i)} - X^{(i)\top} \beta \right)^2 + \lambda \|\beta\|^2$$

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 λ still couples components

Componentwise linear regression: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{C^2\sigma^2}{n} \to \tilde{C}^2 \qquad \frac{d\sigma^2}{n} \to \tilde{d}$$

Componentwise linear regression: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{C^2\sigma^2}{n} \to \tilde{C}^2 \qquad \frac{d\sigma^2}{n} \to \tilde{d}$$

Then:

$$E_n^* \xrightarrow{P} \mathcal{E}^*$$

Componentwise linear regression: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{C^2\sigma^2}{n} \to \tilde{C}^2 \qquad \frac{d\sigma^2}{n} \to \tilde{d}$$

Then:

$$E_n^* = \inf_{\lambda \geq 0} E_n^\lambda \xrightarrow{P} \mathcal{E}^*$$

Componentwise linear regression: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{C^2\sigma^2}{n} \to \tilde{C}^2 \qquad \frac{d\sigma^2}{n} \to \tilde{d}$$

Then:

$$E_n^* = \inf_{\lambda \geq 0} E_n^\lambda \xrightarrow{P} \inf_{\lambda \geq 0} \mathcal{E}^\lambda = \mathcal{E}^*$$

Componentwise linear regression: theorem

If:

$$B^2 \to \tilde{B}^2 \qquad \frac{C^2\sigma^2}{n} \to \tilde{C}^2 \qquad \frac{d\sigma^2}{n} \to \tilde{d}$$

Then:

$$E_n^* = \inf_{\lambda \geq 0} E_n^\lambda \xrightarrow{P} \inf_{\lambda \geq 0} \mathcal{E}^\lambda = \mathcal{E}^*$$

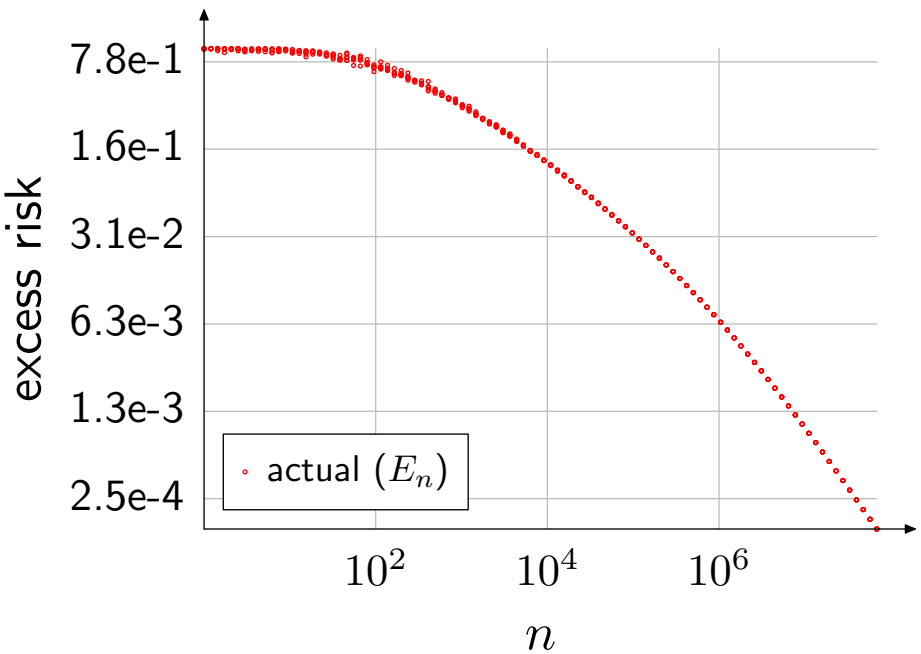
$$\mathcal{E}^\lambda = \underbrace{\frac{\tilde{B}^2\lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$

Componentwise linear regression: intuition

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$

Componentwise linear regression: intuition

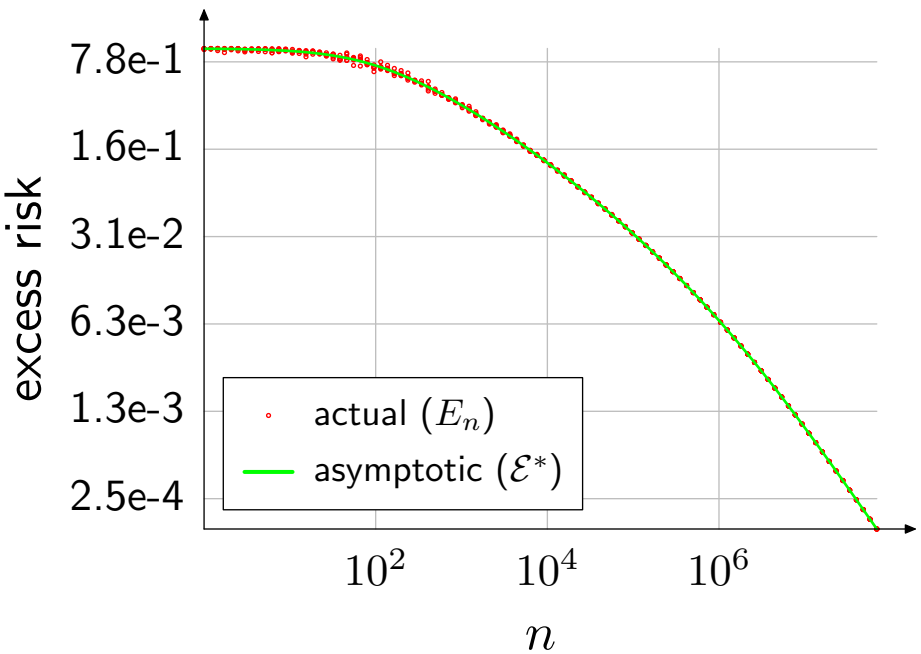
$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$



$B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100$

Componentwise linear regression: intuition

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$

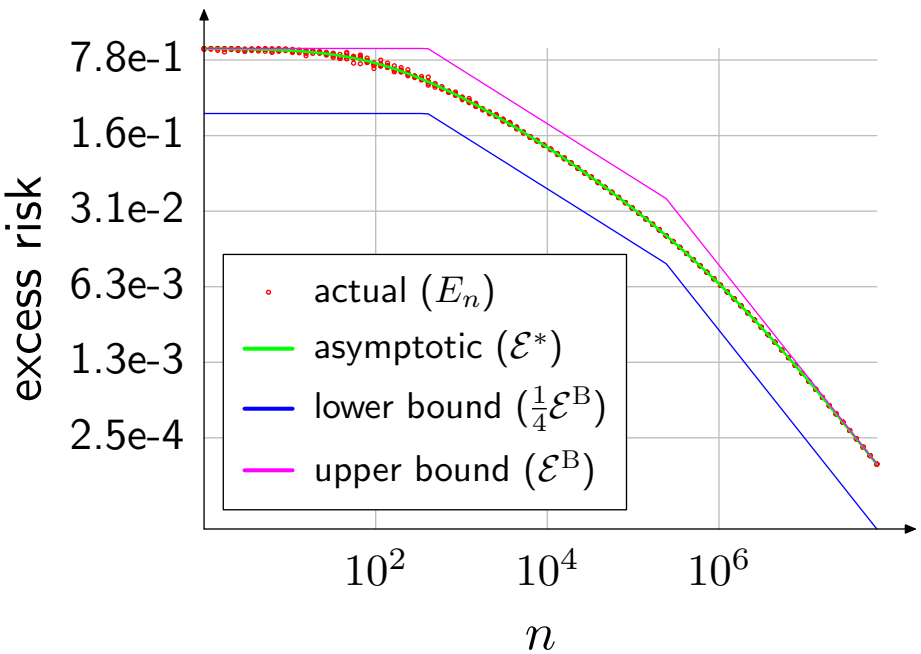


$B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100$

Componentwise linear regression: intuition

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$

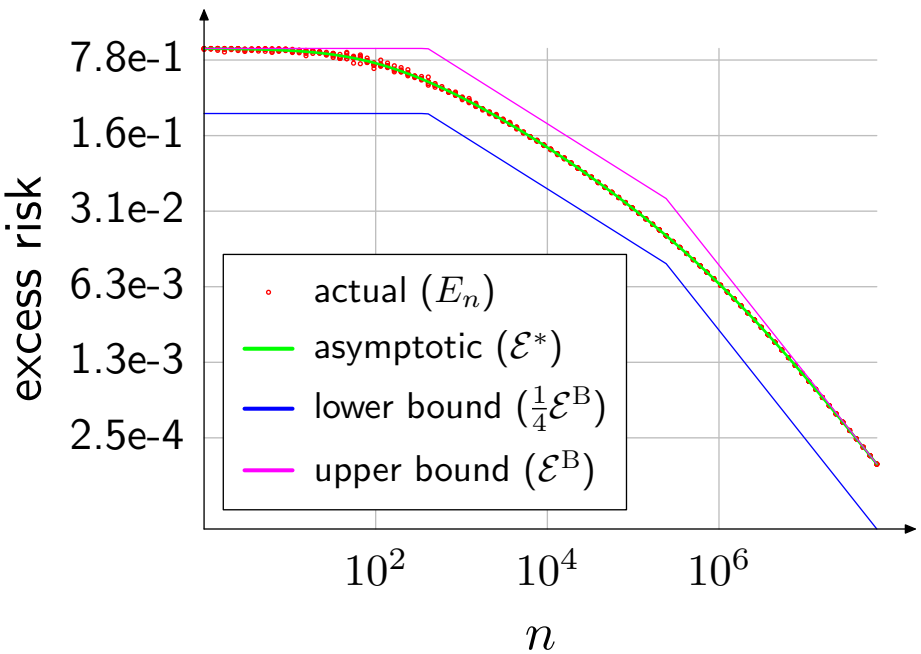
$$\mathcal{E}^{\mathrm{B}} \stackrel{\text{def}}{=} \min \left\{ \tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B}\sqrt{\tilde{d}}}, \tilde{d} \right\}$$



$B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100$

Componentwise linear regression: intuition

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$



$$\mathcal{E}^{\mathrm{B}} \stackrel{\text{def}}{=} \min\left\{\tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B}\sqrt{\tilde{d}}}, \tilde{d}\right\}$$

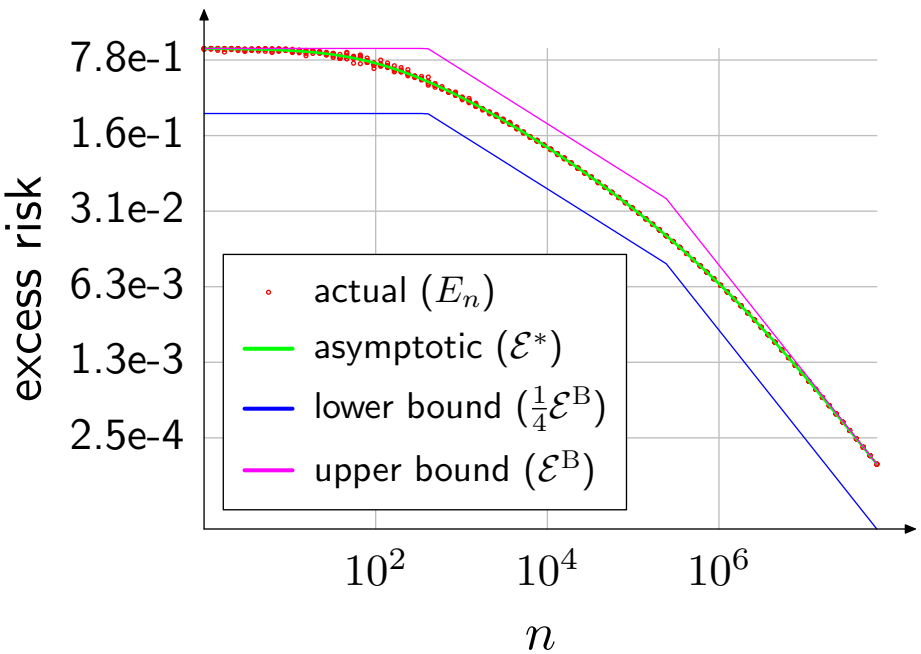
Random ($\lambda \to \infty$):

Bias dominates; $\mathcal{E}^* \sim 1$

$B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100$

Componentwise linear regression: intuition

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$



$$\mathcal{E}^{\mathrm{B}} \stackrel{\text{def}}{=} \min \left\{ \tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B}\sqrt{\tilde{d}}}, \tilde{d} \right\}$$

Random ($\lambda \to \infty$):

Bias dominates; $\mathcal{E}^* \sim 1$

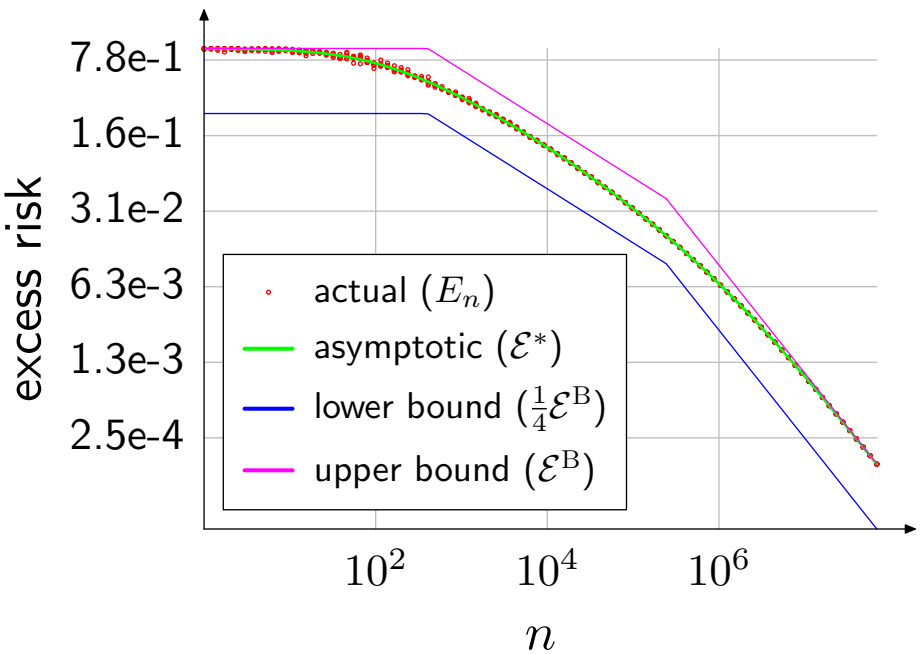
Regularized (λ non-trivial):

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

$$\mathcal{E}^* = \inf_{\lambda \geq 0} \underbrace{\frac{\tilde{B}^2 \lambda^2}{(\tilde{B}^2 + \lambda)^2}}_{\text{squared bias}} + \underbrace{\frac{\frac{\tilde{C}^4}{\tilde{d}}}{(\frac{\tilde{C}^2}{\tilde{d}} + \lambda)^2}}_{\text{variance}}$$



$B^2 = 1, C^2 = 10, d = 100, \sigma^2 = 100$

$$\mathcal{E}^{\mathrm{B}} \stackrel{\text{def}}{=} \min \left\{ \tilde{B}^2, \frac{2\tilde{C}^2}{\tilde{B}\sqrt{\tilde{d}}}, \tilde{d} \right\}$$

Random ($\lambda \to \infty$):

Bias dominates; $\mathcal{E}^* \sim 1$

Regularized (λ non-trivial):

Balance bias/variance; $\mathcal{E}^* \sim \frac{1}{\sqrt{n}}$

Unregularized ($\lambda \to 0$):

Variance dominates; $\mathcal{E}^* \sim \frac{1}{n}$

Full linear regression: speculation

Stitch together results:

$$\mathcal{E}^* \approxeq \min\{ \quad \tilde{B}^2 \quad ,$$

trivial

Full linear regression: speculation

Stitch together results:

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

trivial

bounds from
[Srebro et al., 2010]

Full linear regression: speculation

Stitch together results:

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

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Full linear regression: speculation

Stitch together results:

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

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Four regimes:

Full linear regression: speculation

Stitch together results:

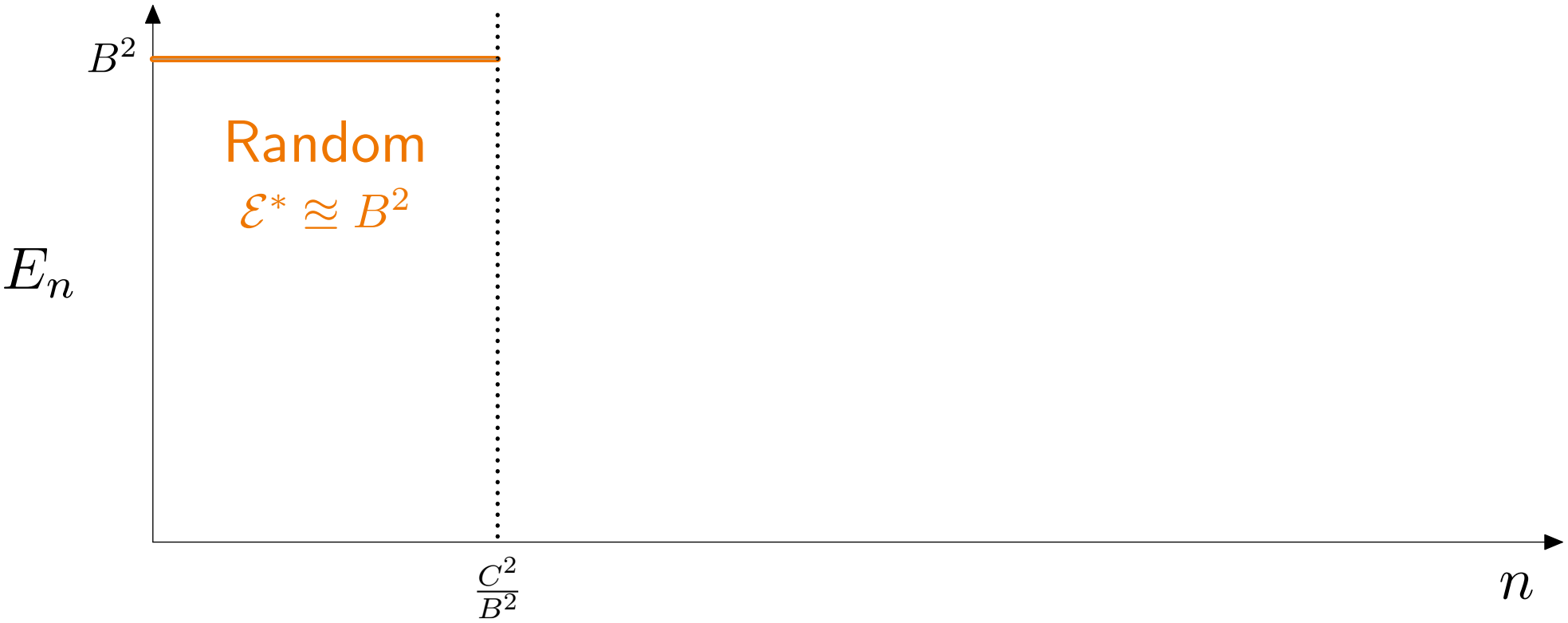
$$\mathcal{E}^* \approxeq \min\{ \quad \tilde{B}^2 \quad , \quad O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right) \quad , \quad \tilde{d} \quad \}$$

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Four regimes:



Full linear regression: speculation

Stitch together results:

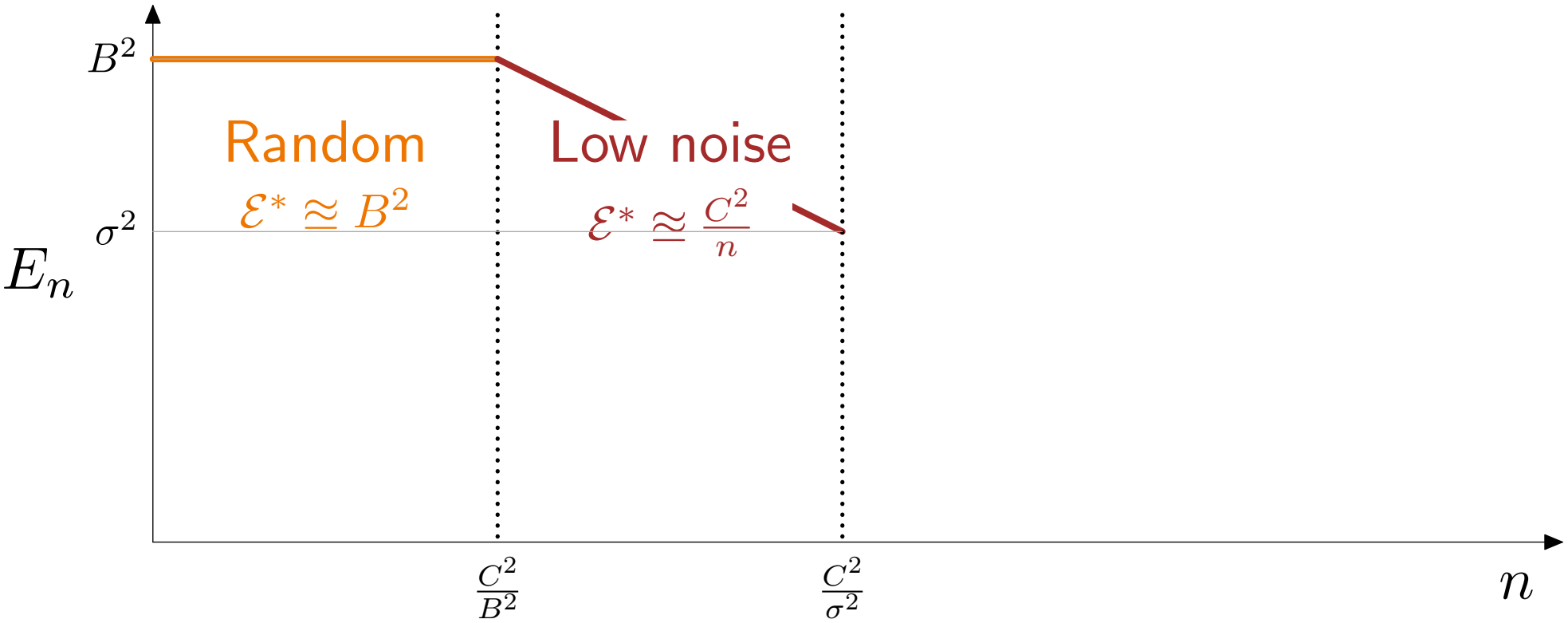
$$\mathcal{E}^* \approxeq \min\{ \quad \tilde{B}^2 \quad , \quad O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right) \quad , \quad \tilde{d} \quad \}$$

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Four regimes:



Full linear regression: speculation

Stitch together results:

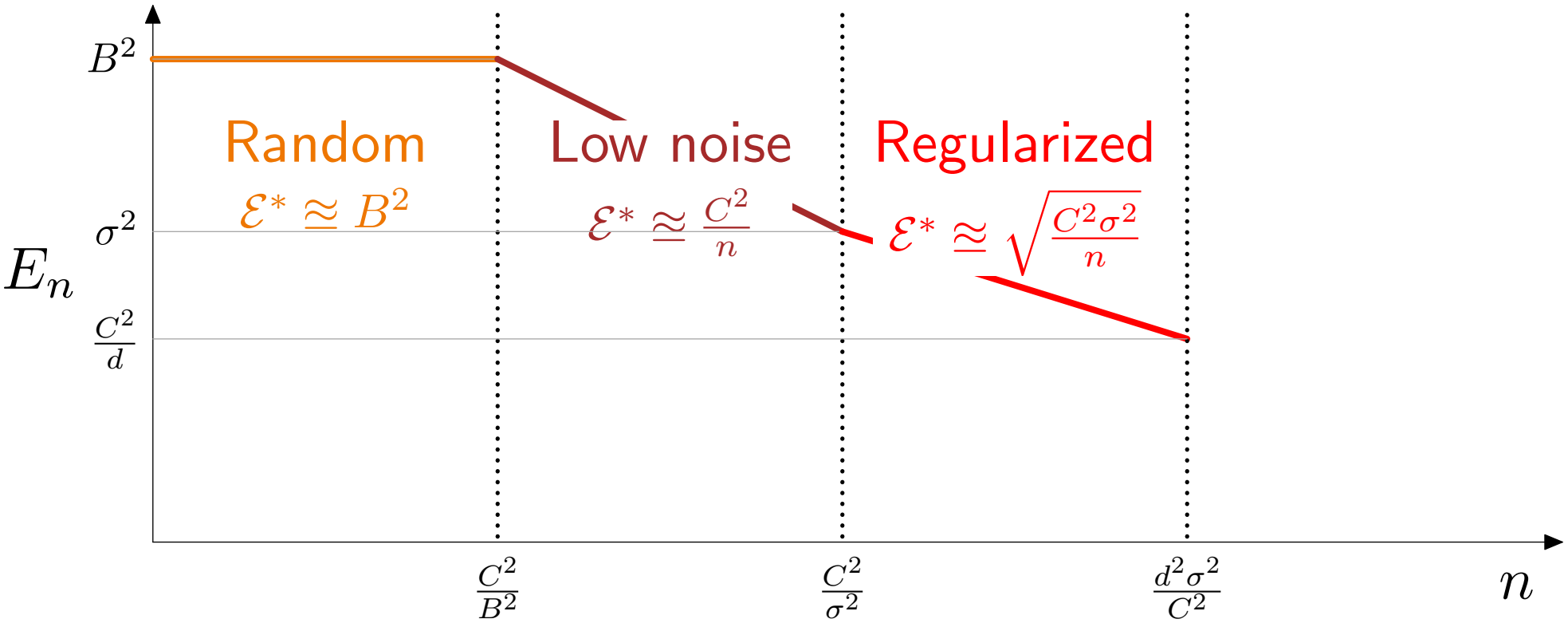
$$\mathcal{E}^* \approxeq \min\{ \quad \tilde{B}^2 \quad , \quad O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right) \quad , \quad \tilde{d} \quad \}$$

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Four regimes:



Full linear regression: speculation

Stitch together results:

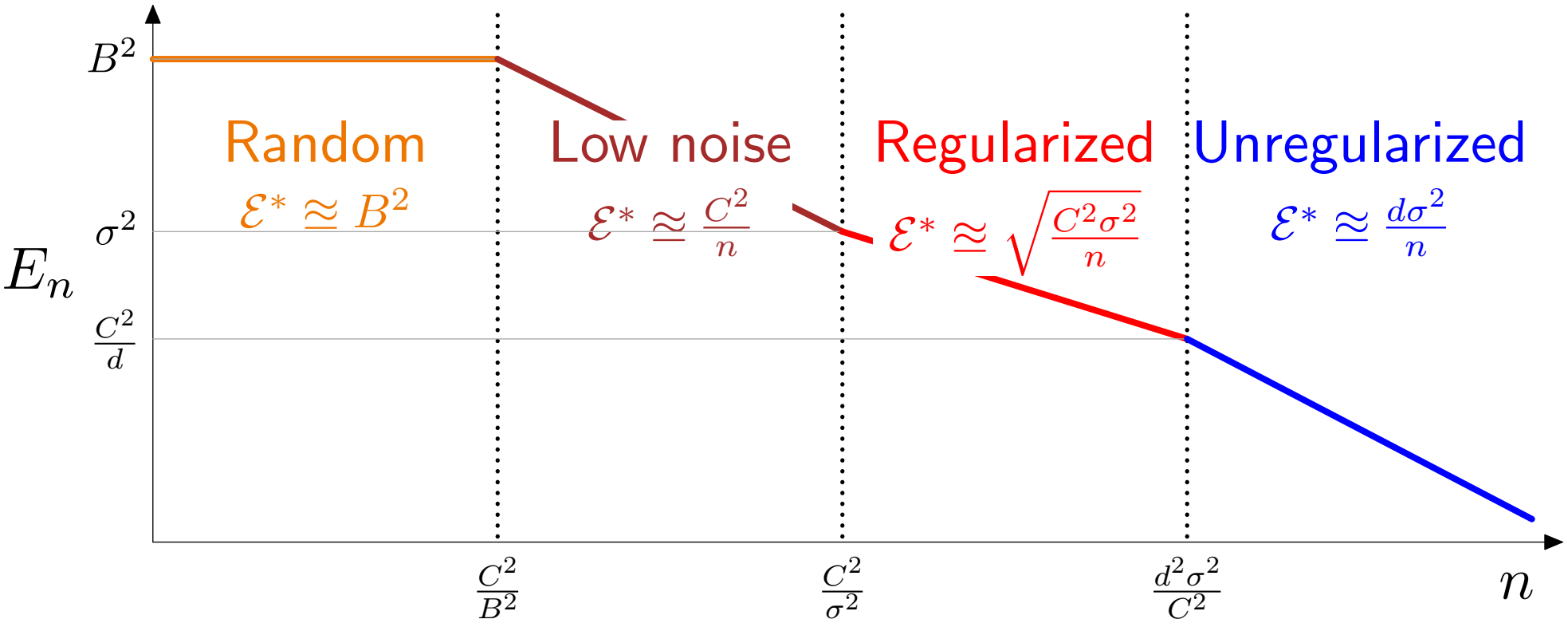
$$\mathcal{E}^* \approxeq \min\{ \quad \tilde{B}^2 \quad , \quad O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right) \quad , \quad \tilde{d} \quad \}$$

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Four regimes:



Full linear regression: speculation

Stitch together results:

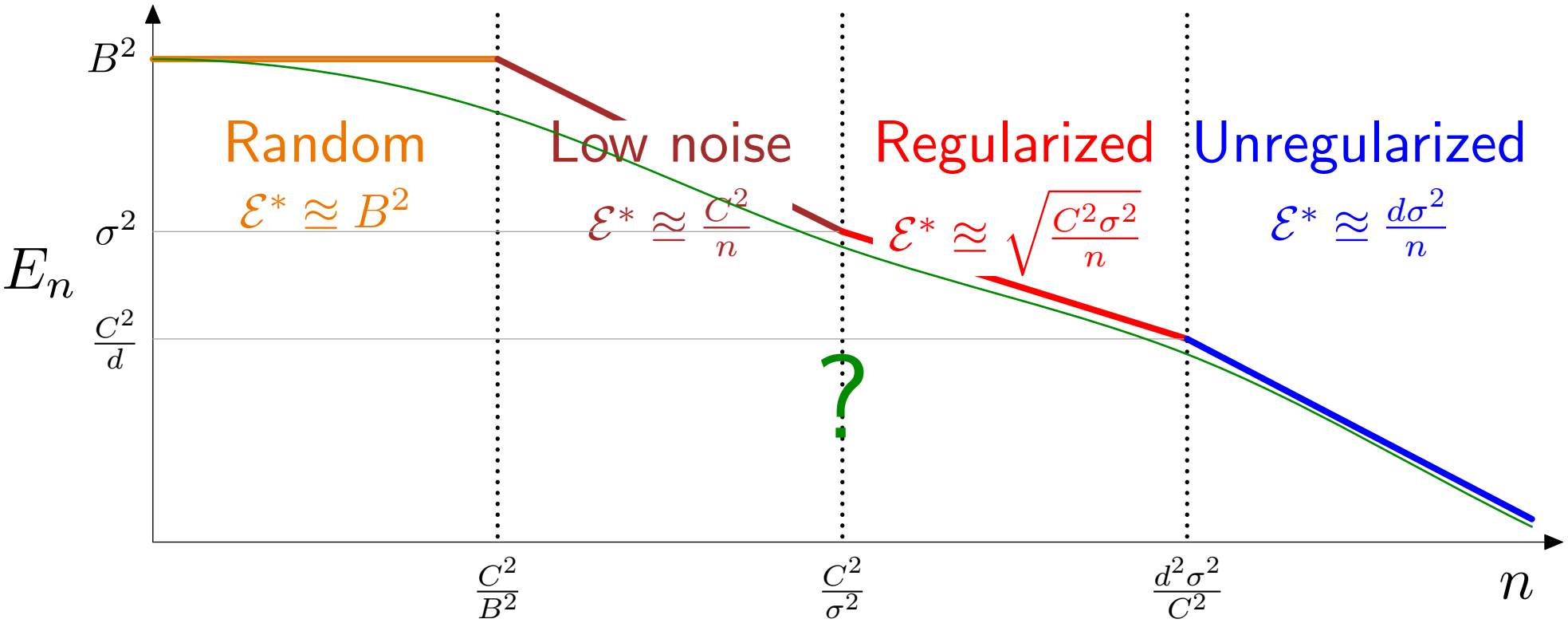
$$\mathcal{E}^* \approxeq \min\{ \quad \tilde{B}^2 \quad , \quad O\left(\frac{\tilde{C}^2}{\tilde{\sigma}^2} + \tilde{C}\right) \quad , \quad \tilde{d} \quad \}$$

trivial

bounds from [Srebro et al., 2010]

classical asymptotics

Four regimes:



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