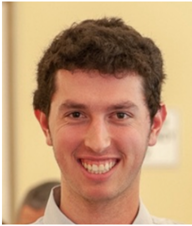
Optimal Mistake Bounds for Transductive Online Learning

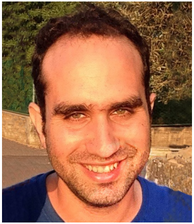
Jonathan Shafer
MIT

November 2025



Zachary Chase

Steve Hanneke



Shay Moran

The Power of Unlabeled Data

Unlabeled Data

Used in:

- Semi-supervised learning
- Self-supervised learning
- ...

Unlabeled Data

Used in:

- Semi-supervised learning
- Self-supervised learning
- ...

Techniques:

- Contrastive learning
- Autoencoders
- Diffusion
- Pretraining, foundation models
- ...

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

A: In **PAC Learning**:

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

A: In **PAC Learning**: the short answer is **NO**.

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

A: In **PAC Learning**: the short answer is **NO**.

Fundamental Theorem: Sample complexity of PAC learning is

$$\Theta\left(\frac{\mathsf{VC} + \log(1/\delta)}{\varepsilon}\right)$$

regardless of whether the distribution on the domain is known.

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

A: In **PAC Learning**: the short answer is **NO**.

Fundamental Theorem: Sample complexity of PAC learning is

$$\Theta\left(\frac{\mathsf{VC} + \log(1/\delta)}{\varepsilon}\right)$$

regardless of whether the distribution on the domain is known.

Q: How about other models of classification?

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

A: In **PAC Learning**: the short answer is **NO**.

Fundamental Theorem: Sample complexity of PAC learning is

$$\Theta\left(\frac{\mathsf{VC} + \log(1/\delta)}{\varepsilon}\right)$$

regardless of whether the distribution on the domain is known.

Q: How about other models of classification?

A: In **Online Learning**:

Back in Theory-Land...

Q: Is unlabeled data useful for classification tasks?

A: In **PAC Learning**: the short answer is **NO**.

Fundamental Theorem: Sample complexity of PAC learning is

$$\Theta\left(\frac{\mathsf{VC} + \log(1/\delta)}{\varepsilon}\right)$$

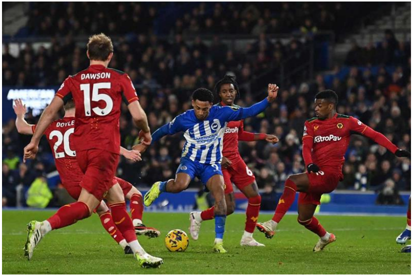
regardless of whether the distribution on the domain is known.

Q: How about other models of classification?

A: In **Online Learning**: **YES**!

Online Learning

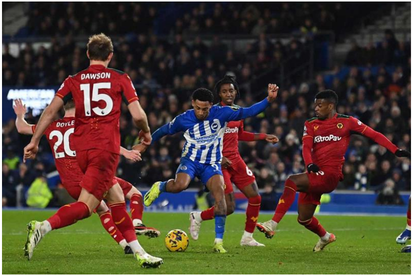
Online Learning: Example



Each day:

1. See what match happens today

Online Learning: Example

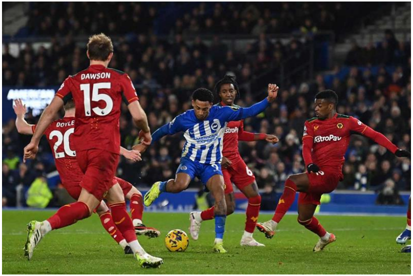


Premier League

Each day:

1. See what match happens today
2. Predict outcome of match

Online Learning: Example



Premier League

Each day:

1. See what match happens today
2. Predict outcome of match
3. Loss / profit

Online Learning: Definition

$\mathcal{X}$ – Domain

$\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ – Hypothesis class

For $t = 1, \ldots, n$:

1. Adversary chooses $x_t \in \mathcal{X}$

Online Learning: Definition

$\mathcal{X}$ – Domain

$\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ – Hypothesis class

For $t = 1, \ldots, n$:

1. Adversary chooses $x_t \in \mathcal{X}$
2. Learner chooses $\hat{y}_t \in \{0,1\}$

Online Learning: Definition

$\mathcal{X}$ – Domain

$\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ – Hypothesis class

For $t = 1, \ldots, n$:

1. Adversary chooses $x_t \in \mathcal{X}$
2. Learner chooses $\hat{y}_t \in \{0,1\}$
3. Adversary chooses $\mathcal{H}$-realizable $y_t \in \{0,1\}$

Online Learning: Definition

$\mathcal{X}$ – Domain

$\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ – Hypothesis class

For $t = 1, \ldots, n$:

1. Adversary chooses $x_t \in \mathcal{X}$
2. Learner chooses $\hat{y}_t \in \{0,1\}$
3. Adversary chooses $\mathcal{H}$-realizable $y_t \in \{0,1\}$
4. Learner pays $\mathbb{1}(\hat{y}_t \neq y_t)$

Online Learning: Definition

$\mathcal{X}$ – Domain

$\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ – Hypothesis class

For $t = 1, \ldots, n$:

1. Adversary chooses $x_t \in \mathcal{X}$
2. Learner chooses $\hat{y}_t \in \{0,1\}$
3. Adversary chooses $\mathcal{H}$-realizable $y_t \in \{0,1\}$
4. Learner pays $\mathbb{1}(\hat{y}_t \neq y_t)$

Want to minimize loss

This Talk: Transductive Online Learning

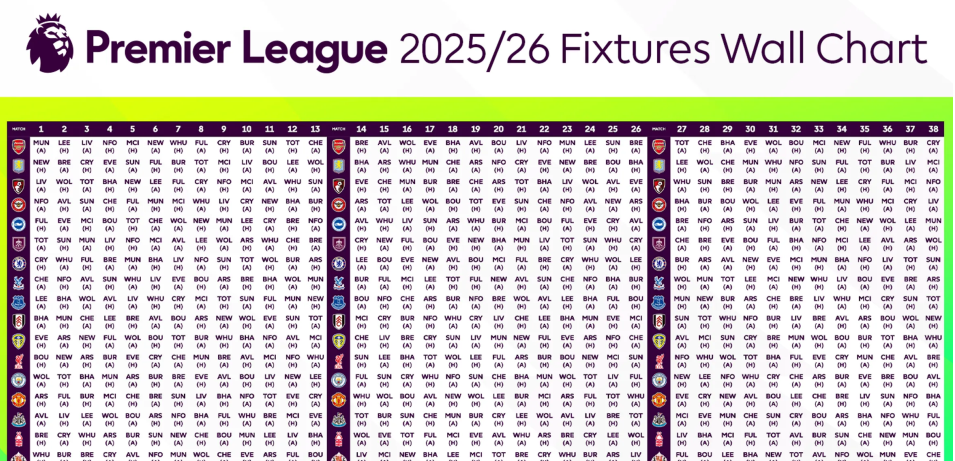
Example:

Look up schedule for entire Premier League season in advance

This Talk: Transductive Online Learning

Example:

Look up schedule for entire Premier League season in advance



Premier League 2025/26 Fixtures Wall Chart

MATCH	1	2	3	4	5	6	7	8	9	10	11	12	13
	MUN (A)	LEE (H)	LIV (A)	NFO (H)	MCI (H)	NEW (A)	WHU (H)	FUL (A)	CRY (H)	BUR (A)	SUN (A)	TOT (H)	CHE (A)
	NEW (H)	BRE (A)	CRY (H)	EVE (A)	SUN (A)	FUL (H)	BUR (H)	TOT (A)	MCI (H)	LIV (A)	BOU (H)	LEE (A)	WOL (H)
	LIV (A)	WOL (H)	TOT (A)	BHA (H)	NEW (H)	LEE (A)	FUL (H)	CRY (A)	NFO (H)	MCI (A)	AVL (A)	WHU (H)	SUN (A)
	NFO (A)	AVL (H)	SUN (A)	CHE (H)	FUL (A)	MUN (H)	MCI (H)	WHU (A)	LIV (H)	CRY (A)	NEW (H)	BHA (A)	BUR (H)
	FUL (H)	EVE (A)	MCI (H)	BOU (A)	TOT (H)	CHE (A)	WOL (A)	NEW (H)	MUN (A)	LEE (H)	CRY (A)	BRE (H)	NFO (A)
	TOT (A)	SUN (H)	MUN (A)	LIV (H)	NFO (H)	MCI (A)	AVL (A)	LEE (H)	WOL (A)	ARS (H)	WHU (A)	CHE (H)	BRE (A)
	CRY (H)	WHU (A)	FUL (H)	BRE (A)	MUN (A)	BHA (H)	LIV (H)	NFO (A)	SUN (H)	TOT (A)	WOL (H)	BUR (A)	ARS (H)
	CHE (A)	NFO (H)	AVL (A)	SUN (H)	WHU (A)	LIV (H)	EVE (A)	BOU (H)	ARS (A)	BRE (H)	BHA (H)	WOL (A)	MUN (H)
	LEE (A)	BHA (H)	WOL (A)	AVL (H)	LIV (A)	WHU (H)	CRY (H)	MCI (A)	TOT (H)	SUN (A)	FUL (H)	MUN (A)	NEW (H)
	BHA (A)	MUN (H)	CHE (A)	LEE (H)	BRE (H)	AVL (A)	BOU (A)	ARS (H)	NEW (A)	WOL (H)	EVE (A)	SUN (H)	TOT (A)
	EVE (H)	ARS (A)	NEW (H)	FUL (A)	WOL (A)	BOU (H)	TOT (H)	BUR (A)	WHU (H)	BHA (A)	NFO (A)	AVL (H)	MCI (A)
	BOU (H)	NEW (A)	ARS (H)	BUR (A)	EVE (H)	CRY (A)	CHE (A)	MUN (H)	BRE (A)	AVL (H)	MCI (A)	NFO (H)	WHU (A)
	WOL (A)	TOT (H)	BHA (A)	MUN (H)	ARS (A)	BUR (H)	BRE (A)	EVE (H)	AVL (A)	BOU (H)	LIV (H)	NEW (A)	LEE (H)
	ARS (H)	FUL (A)	BUR (H)	MCI (A)	CHE (H)	BRE (A)	SUN (H)	LIV (A)	BHA (H)	NFO (A)	TOT (A)	EVE (H)	CRY (A)
	AVL (A)	LIV (H)	LEE (A)	WOL (H)	BOU (A)	ARS (H)	NFO (H)	BHA (A)	FUL (H)	WHU (A)	BRE (A)	MCI (H)	EVE (A)
	BRE (H)	CRY (A)	WHU (H)	ARS (A)	BUR (A)	SUN (H)	NEW (A)	CHE (H)	BOU (A)	MUN (H)	LEE (H)	LIV (A)	BHA (H)
	WHU	BUR	BRE	CRY	AVL	NFO	MUN	WOL	CHE	EVE	ARS	FUL	BOU

MATCH	14	15	16	17	18	19	20	21	22	23	24	25	26
	BRE (H)	AVL (A)	WOL (H)	EVE (A)	BHA (H)	AVL (H)	BOU (A)	LIV (H)	NFO (A)	MUN (H)	LEE (A)	SUN (H)	BRE (A)
	BHA (A)	ARS (H)	WHU (A)	MUN (H)	CHE (A)	ARS (A)	NFO (H)	CRY (A)	EVE (H)	NEW (A)	BRE (H)	BOU (A)	BHA (H)
	EVE (H)	CHE (H)	MUN (A)	BUR (H)	BRE (A)	CHE (A)	ARS (H)	TOT (H)	BHA (A)	LIV (H)	WOL (A)	AVL (H)	EVE (A)
	ARS (A)	TOT (A)	LEE (H)	WOL (A)	BOU (H)	TOT (H)	EVE (A)	SUN (H)	CHE (A)	NFO (H)	AVL (A)	NEW (A)	ARS (H)
	AVL (H)	WHU (H)	LIV (A)	SUN (H)	ARS (A)	WHU (A)	BUR (H)	MCI (A)	BOU (H)	FUL (A)	EVE (H)	CRY (H)	AVL (A)
	CRY (H)	NEW (A)	FUL (H)	BOU (A)	EVE (H)	NEW (H)	BHA (A)	MUN (H)	LIV (A)	TOT (H)	SUN (A)	WHU (H)	CRY (A)
	LEE (A)	BOU (A)	EVE (H)	NEW (A)	AVL (H)	BOU (H)	MCI (A)	FUL (A)	BRE (H)	CRY (A)	WHU (H)	WOL (A)	LEE (H)
	BUR (A)	FUL (A)	MCI (H)	LEE (A)	TOT (H)	FUL (H)	NEW (A)	AVL (H)	SUN (A)	CHE (H)	NFO (A)	BHA (A)	BUR (H)
	BOU (A)	NFO (H)	CHE (A)	ARS (H)	BUR (A)	NFO (A)	BRE (H)	WOL (H)	AVL (A)	LEE (H)	BHA (A)	FUL (A)	BOU (H)
	MCI (H)	CRY (H)	BUR (A)	NFO (H)	WHU (A)	CRY (A)	LIV (H)	CHE (H)	LEE (A)	BHA (H)	MUN (A)	EVE (H)	MCI (A)
	CHE (H)	LIV (H)	BRE (A)	CRY (H)	SUN (A)	LIV (A)	MUN (H)	NEW (A)	FUL (H)	EVE (A)	ARS (H)	NFO (H)	CHE (A)
	SUN (H)	LEE (A)	BHA (H)	TOT (A)	WOL (H)	LEE (H)	FUL (A)	ARS (A)	BUR (H)	BOU (A)	NEW (H)	MCI (H)	SUN (A)
	FUL (A)	SUN (H)	CRY (A)	WHU (H)	NFO (A)	SUN (A)	CHE (H)	BHA (H)	MUN (A)	WOL (H)	TOT (A)	LIV (A)	FUL (H)
	WHU (H)	WOL (A)	BOU (H)	AVL (A)	NEW (H)	WOL (H)	LEE (A)	BUR (A)	MCI (H)	ARS (A)	FUL (H)	TOT (H)	WHU (A)
	TOT (H)	BUR (H)	SUN (A)	CHE (H)	MUN (A)	BUR (A)	CRY (H)	LEE (H)	WOL (A)	AVL (H)	LIV (A)	BRE (H)	TOT (A)
	WOL (A)	EVE (A)	TOT (H)	FUL (A)	MCI (H)	EVE (H)	AVL (A)	WHU (A)	ARS (H)	BRE (A)	CRY (H)	LEE (A)	WOL (H)
	LIV	MCI	NEW	BHA	LEE	MCI	TOT	BRE	CRY	WHU	BUR	ARS	LIV

MATCH	27	28	29	30	31	32	33	34	35	36	37	38
	TOT (A)	CHE (H)	BHA (A)	EVE (H)	WOL (A)	BOU (H)	MCI (A)	NEW (H)	FUL (A)	WHU (A)	BUR (H)	CRY (A)
	LEE (H)	WOL (A)	CHE (H)	MUN (A)	WHU (H)	NFO (A)	SUN (H)	FUL (A)	TOT (H)	BUR (A)	LIV (H)	MCI (A)
	WHU (A)	SUN (H)	BRE (H)	BUR (A)	MUN (H)	ARS (A)	NEW (A)	LEE (H)	CRY (H)	FUL (A)	MCI (H)	NFO (A)
	BHA (H)	BUR (A)	BOU (A)	WOL (H)	LEE (A)	EVE (H)	FUL (H)	MUN (A)	WHU (H)	MCI (A)	CRY (H)	LIV (A)
	BRE (A)	NFO (H)	ARS (H)	SUN (A)	LIV (A)	BUR (H)	TOT (A)	CHE (H)	NEW (A)	WOL (H)	LEE (A)	MUN (H)
	CHE (H)	BRE (H)	EVE (A)	BOU (H)	FUL (A)	BHA (A)	NFO (A)	MCI (H)	LEE (H)	AVL (H)	ARS (A)	WOL (H)
	BUR (A)	ARS (A)	AVL (A)	NEW (H)	EVE (A)	MCI (H)	MUN (H)	BHA (A)	NFO (H)	LIV (A)	TOT (H)	SUN (A)
	WOL (H)	MUN (A)	TOT (A)	LEE (H)	MCI (A)	NEW (H)	WHU (H)	LIV (A)	BOU (A)	EVE (H)	BRE (A)	ARS (H)
	MUN (H)	NEW (A)	BUR (H)	ARS (A)	CHE (H)	BRE (A)	LIV (H)	WHU (A)	MCI (H)	CRY (A)	SUN (H)	TOT (A)
	SUN (A)	TOT (H)	WHU (H)	NFO (A)	BUR (H)	LIV (A)	BRE (A)	AVL (H)	ARS (H)	BOU (A)	WOL (H)	NEW (H)
	AVL (A)	MCI (H)	SUN (H)	CRY (A)	BRE (H)	MUN (A)	WOL (H)	BOU (A)	BUR (A)	TOT (H)	BHA (A)	WHU (H)
	NFO (H)	WHU (H)	WOL (A)	TOT (A)	BHA (H)	FUL (H)	EVE (A)	CRY (H)	MUN (A)	CHE (H)	AVL (A)	BRE (H)
	NEW (H)	LEE (A)	NFO (H)	WHU (A)	CRY (H)	CHE (A)	ARS (H)	BUR (A)	EVE (A)	BRE (H)	BOU (A)	AVL (H)
	EVE (A)	CRY (H)	NEW (A)	AVL (H)	BOU (A)	LEE (H)	CHE (A)	BRE (H)	LIV (H)	SUN (A)	NFO (H)	BHA (A)
	MCI (A)	EVE (H)	MUN (H)	CHE (A)	SUN (H)	CRY (A)	BOU (H)	ARS (A)	BHA (H)	NFO (A)	WHU (H)	FUL (A)
	LIV (H)	BHA (A)	MCI (A)	FUL (H)	TOT (A)	AVL (H)	BUR (H)	SUN (A)	CHE (A)	NEW (H)	MUN (A)	BOU (H)
	FUL	BOU	LEE	BHA	NEW	TOT	AVL	NFO	WOL	MUN	EVE	CHE

This Talk: Transductive Online Learning

Example:

Look up schedule for entire Premier League season in advance

Adversary chooses $x_1, \ldots, x_n \in \mathcal{X}$

This Talk: Transductive Online Learning

Example:

Look up schedule for entire Premier League season in advance

Adversary chooses $x_1, \ldots, x_n \in \mathcal{X}$

For $t = 1, \ldots, n$:

1. Learner chooses $\hat{y}_t \in \{0, 1\}$

This Talk: Transductive Online Learning

Example:

Look up schedule for entire Premier League season in advance

Adversary chooses $x_1, \ldots, x_n \in \mathcal{X}$

For $t = 1, \ldots, n$:

1. Learner chooses $\hat{y}_t \in \{0, 1\}$
2. Adversary chooses $\mathcal{H}$-realizable $y_t \in \{0, 1\}$

Definition: Ben-David, Kushilevitz and Mansour (1995)

This Talk: Transductive Online Learning

Example:

Look up schedule for entire Premier League season in advance

Adversary chooses $x_1, \ldots, x_n \in \mathcal{X}$

For $t = 1, \ldots, n$:

1. Learner chooses $\hat{y}_t \in \{0, 1\}$
2. Adversary chooses $\mathcal{H}$-realizable $y_t \in \{0, 1\}$
3. Learner pays $\mathbb{1}(\hat{y}_t \neq y_t)$

Definition: Ben-David, Kushilevitz and Mansour (1995)

Main Question

Transductive online learning:

Learner has access to

full sequence of unlabeled instances in advance

Main Question

Transductive online learning:

Learner has access to

full sequence of unlabeled instances in advance

Q: How much does access to unlabeled data help?

Q: How much does access to unlabeled data help?

Q: How much does access to unlabeled data help?

Lower Bound **Upper Bound**

Q: How much does access to unlabeled data help?

Lower Bound **Upper Bound**

$$M_{\text{tr}} \leq \qquad M_{\text{std}}$$

Q: How much does access to unlabeled data help?

Lower Bound		**Upper Bound**	
	$M_{\text{tr}} \leq$	M_{std}	
$\Omega(\log\log M_{\text{std}})$	$\leq M_{\text{tr}} \leq$	$\frac{2}{3} \cdot M_{\text{std}}$	[BKM95]

Q: How much does access to unlabeled data help?

Lower Bound		Upper Bound	
	$M_{\text{tr}} \leq$	M_{std}	
$\Omega(\log\log M_{\text{std}})$	$\leq M_{\text{tr}} \leq$	$\frac{2}{3} \cdot M_{\text{std}}$	[BKM95]
$\Omega\left(\sqrt{\log M_{\text{std}}}\right)$	$\leq M_{\text{tr}}$		[BKM97]

Q: How much does access to unlabeled data help?

Lower Bound		Upper Bound	
	$M_{\mathrm{tr}} \le$	M_{std}	
$\Omega(\log\log M_{\mathrm{std}})$	$\le M_{\mathrm{tr}} \le$	$\frac{2}{3} \cdot M_{\mathrm{std}}$	[BKM95]
$\Omega\left(\sqrt{\log M_{\mathrm{std}}}\right)$	$\le M_{\mathrm{tr}}$		[BKM97]
$\Omega(\log M_{\mathrm{std}})$	$\le M_{\mathrm{tr}}$		[HMS23]

Q: How much does access to unlabeled data help?

Lower Bound		Upper Bound	
	$M_{\text{tr}} \le$	M_{std}	
$\Omega(\log\log M_{\text{std}})$	$\le M_{\text{tr}} \le$	$\frac{2}{3} \cdot M_{\text{std}}$	[BKM95]
$\Omega\left(\sqrt{\log M_{\text{std}}}\right)$	$\le M_{\text{tr}}$		[BKM97]
$\Omega(\log M_{\text{std}})$	$\le M_{\text{tr}}$		[HMS23]
$\Omega\left(\sqrt{M_{\text{std}}}\right)$	$\le M_{\text{tr}} \le$	$O\left(\sqrt{M_{\text{std}}}\right)$	This Work

Main Result

Q: How much does access to unlabeled data help?

Main Result

Q: How much does access to unlabeled data help?

A:

Theorem.

Main Result

Q: How much does access to unlabeled data help?

A:

Theorem.

For every class $\mathcal{H}$:

Main Result

Q: How much does access to unlabeled data help?

A:

Theorem.

For every class $\mathcal{H}$:

$$M_{\mathrm{tr}}(\mathcal{H}) \geq \Omega\left(\sqrt{M_{\mathrm{std}}(\mathcal{H})}\right).$$

Main Result

Q: How much does access to unlabeled data help?

A:

Theorem.

For every class $\mathcal{H}$:

$$M_{\mathsf{tr}}(\mathcal{H}) \geq \Omega\left(\sqrt{M_{\mathsf{std}}(\mathcal{H})}\right).$$

For all d there exists $\mathcal{H}$:

Main Result

Q: How much does access to unlabeled data help?

A:

Theorem.

For every class $\mathcal{H}$:

$$M_{\mathrm{tr}}(\mathcal{H}) \geq \Omega\left(\sqrt{M_{\mathrm{std}}(\mathcal{H})}\right).$$

For all d there exists $\mathcal{H}$:

$M_{\mathrm{std}}(\mathcal{H}) = d$, and

$$M_{\mathrm{tr}}(\mathcal{H}) \leq O\left(\sqrt{M_{\mathrm{std}}(\mathcal{H})}\right).$$

Techniques

Danger zone minimization

Techniques

Danger zone minimization

Probabilistic construction

Techniques

Danger zone minimization

Probabilistic construction

Sparse codewords are easy to guess

Techniques

Danger zone minimization

Probabilistic construction

Sparse codewords are easy to guess

Splitting experts

Techniques

Danger zone minimization

Probabilistic construction

Sparse codewords are easy to guess

Splitting experts

...

Thank You!

References

[BKM95] Shai Ben-David, Eyal Kushilevitz, and Yishay Mansour. Online learning versus offline learning. In Paul M. B. Vitányi, editor, *Computational Learning Theory, Second European Conference, EuroCOLT '95, Barcelona, Spain, March 13-15, 1995, Proceedings*, volume 904 of *Lecture Notes in Computer Science*, pages 38–52. Springer, 1995.

[BKM97] Shai Ben-David, Eyal Kushilevitz, and Yishay Mansour. Online learning versus offline learning. *Mach. Learn.*, 29(1):45–63, 1997.

[HMS23] Steve Hanneke, Shay Moran, and Jonathan Shafer. A trichotomy for transductive online learning. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.