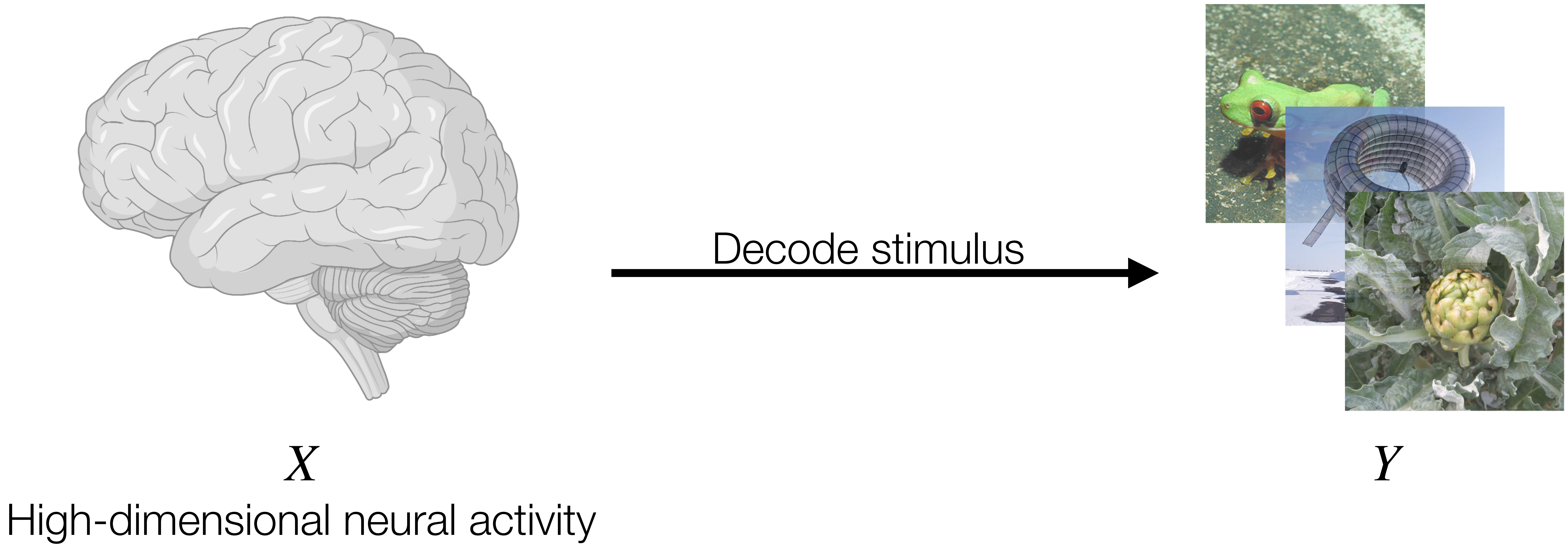
Factor-Analytic Inverse Regression for High-dimensional, Small-sample Dimensionality Reduction

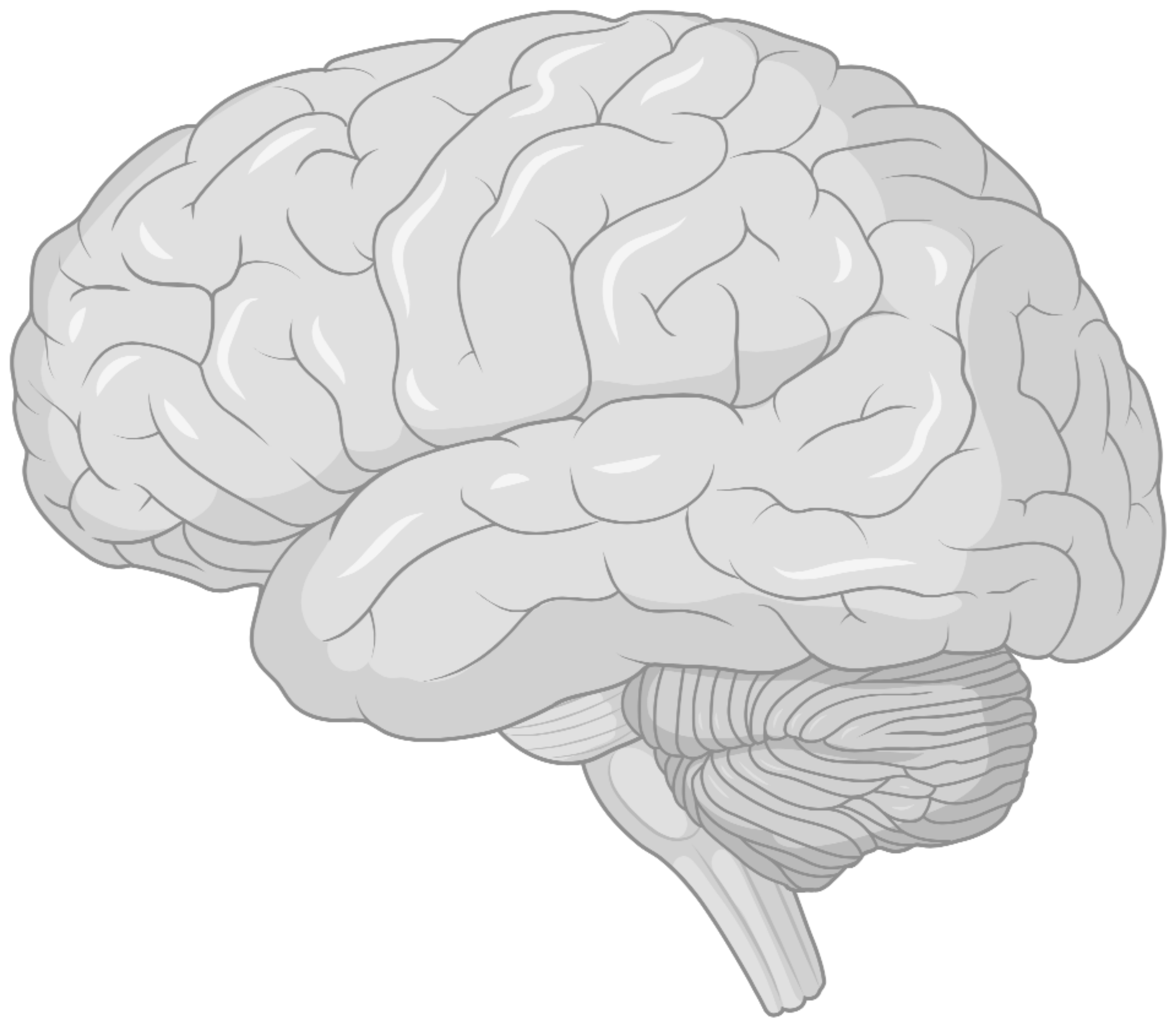
Aditi Jha*, Michael J. Morais*, Jonathan W. Pillow
Princeton Neuroscience Institute

X

High-dimensional neural activity


Decode stimulus
X
High-dimensional neural activity
Y

Can we find a subset of dimensions which preserves information about Y in X?



Decode stimulus

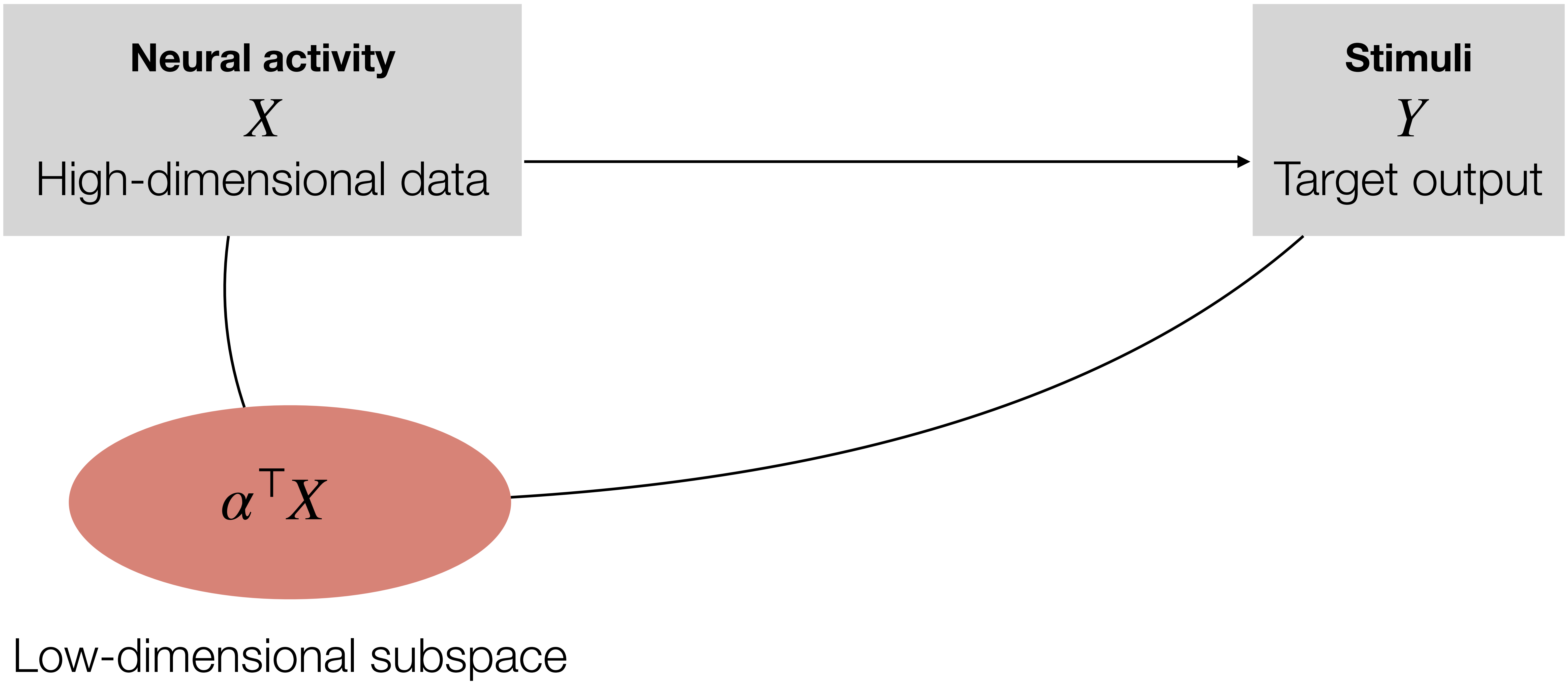
X

High-dimensional neural activity

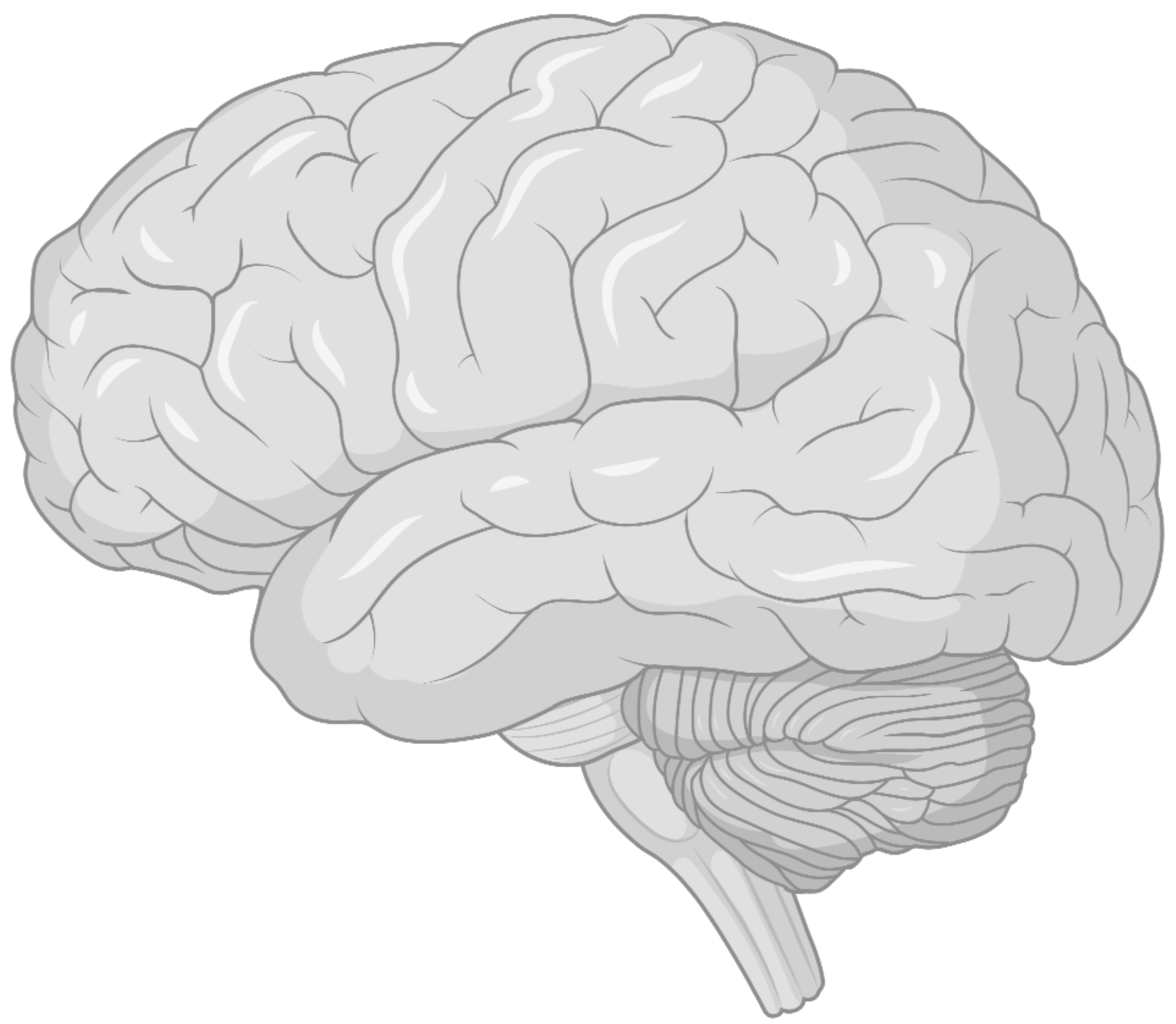
Y

Sufficient Dimension Reduction (SDR)

Can we find a subset of dimensions which preserves information about Y in X?



Can we find a subset of dimensions which preserves information about Y in X?



Decode stimulus

X

High-dimensional neural activity

Y

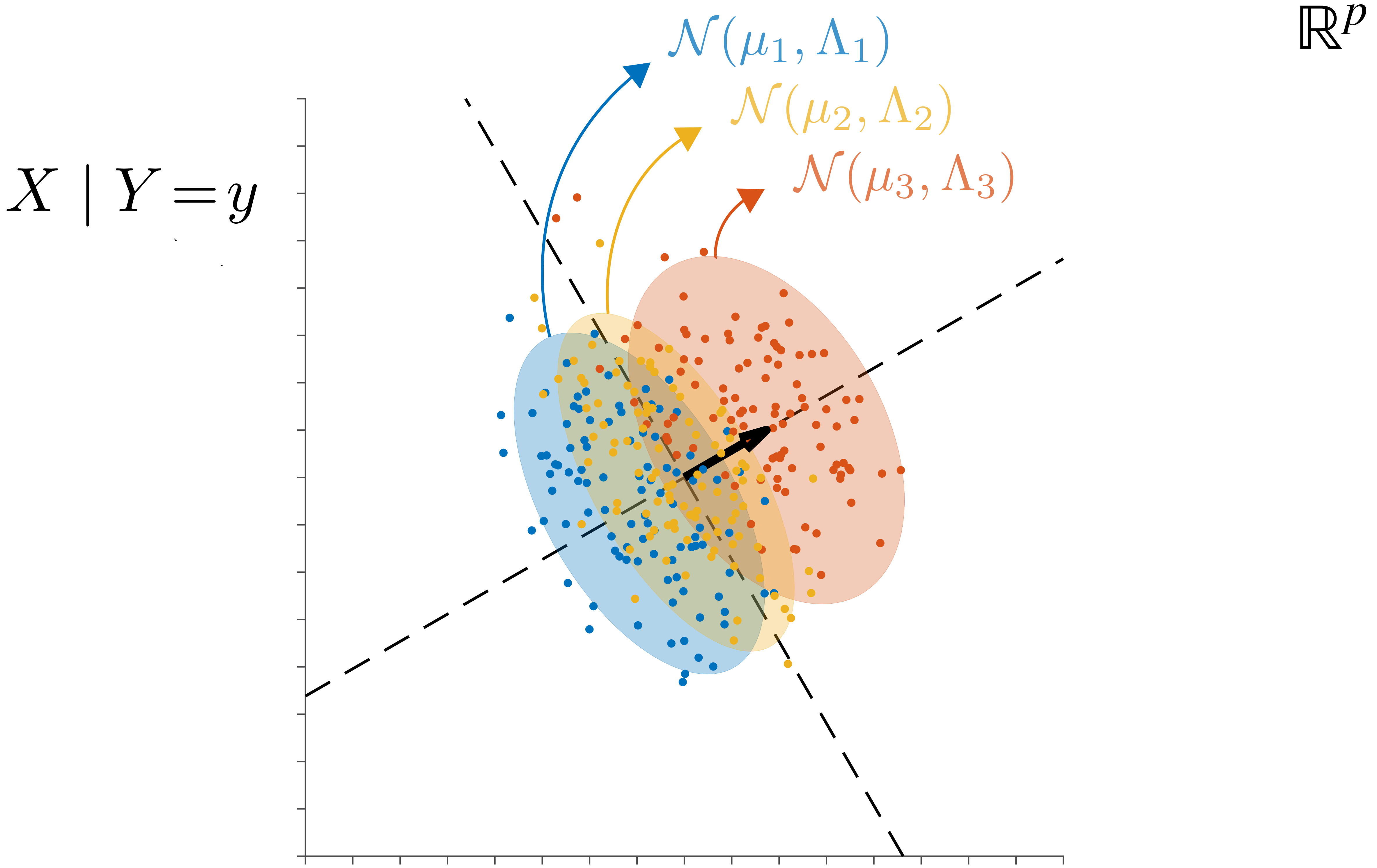
Data is precious!

High-dimensional, small-sample size

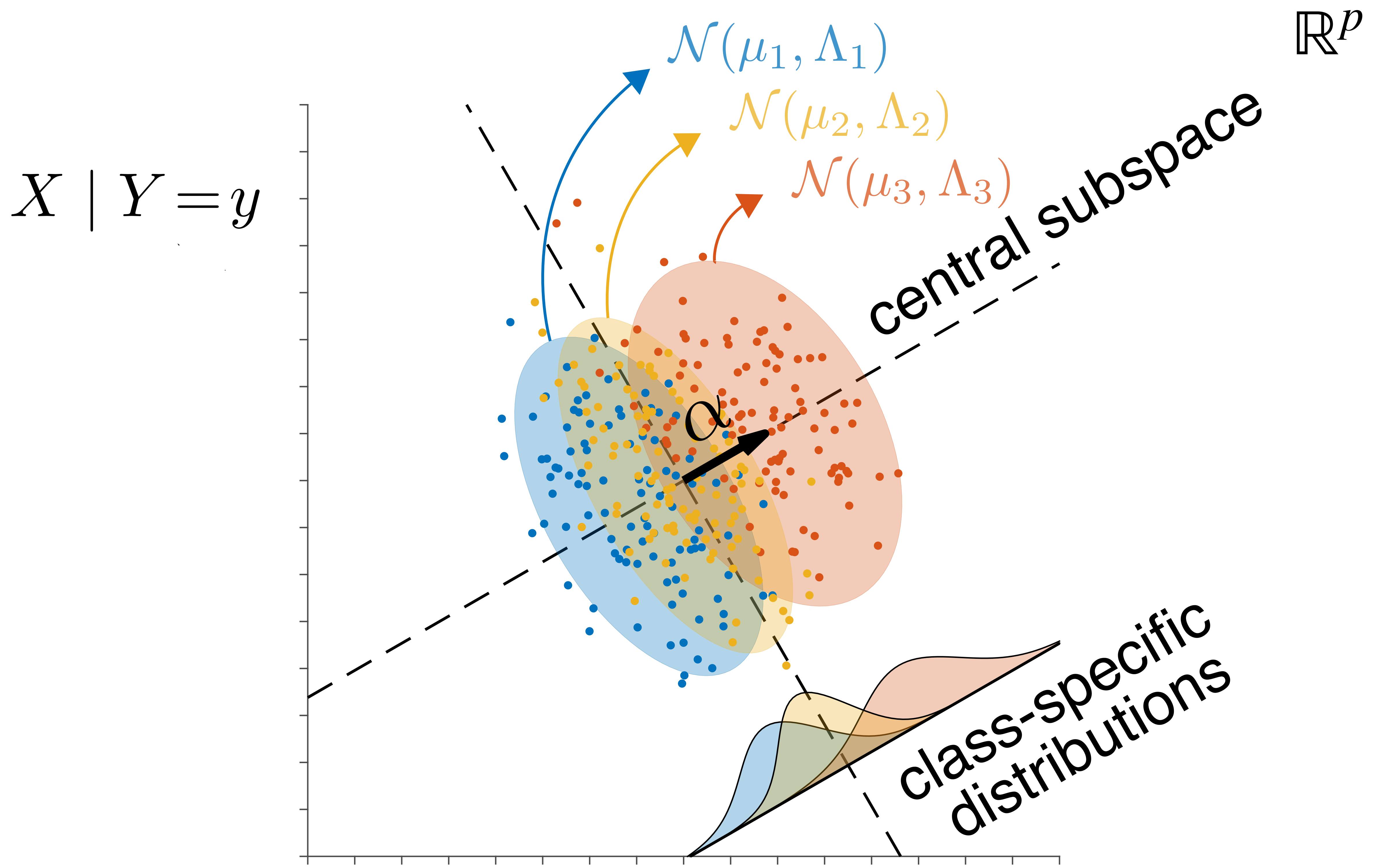
Sufficient Dimension Reduction (SDR)

- N (number of samples) $> p$ (data-dimensionality)
- Performance degrades sharply with decrease in N

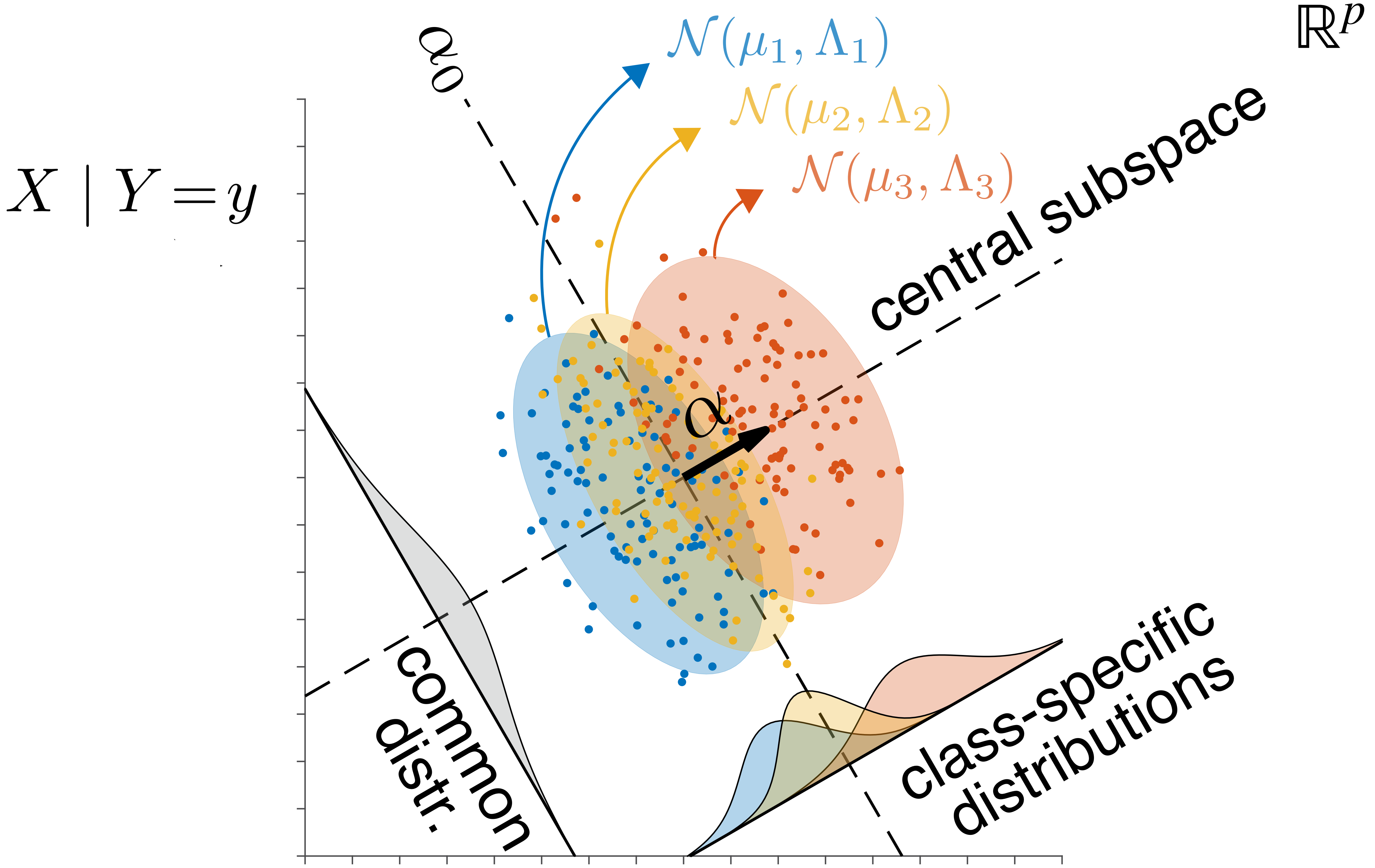
Class-Conditional Factor Analytic Dimensions (CFAD)



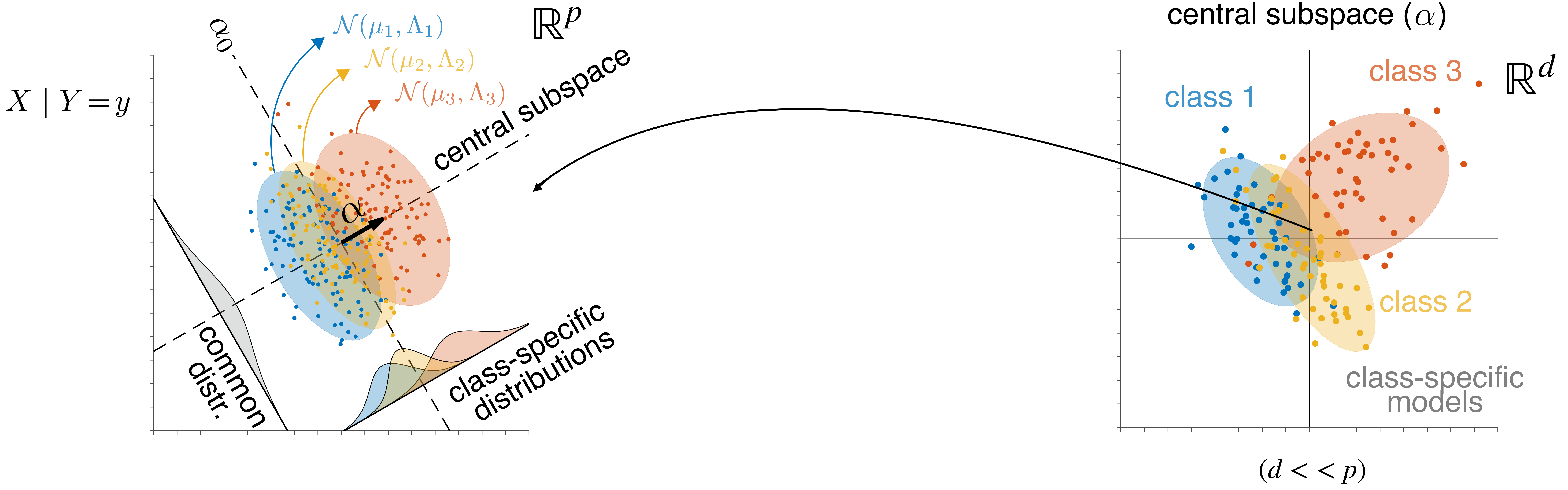
Class-Conditional Factor Analytic Dimensions (CFAD)



Class-Conditional Factor Analytic Dimensions (CFAD)



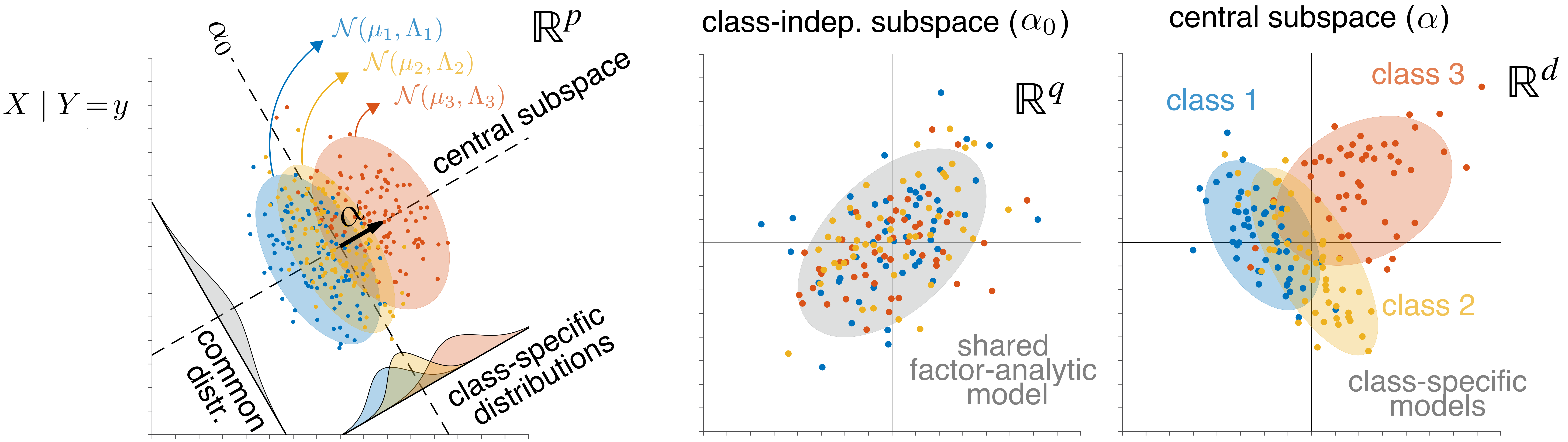
Class-Conditional Factor Analytic Dimensions (CFAD)



$$X_{CS} \mid Y \sim \mathcal{N}(\alpha\nu_y, \alpha\Sigma_y\alpha^\top)$$

Class-specific
covariance

Class-Conditional Factor Analytic Dimensions (CFAD)

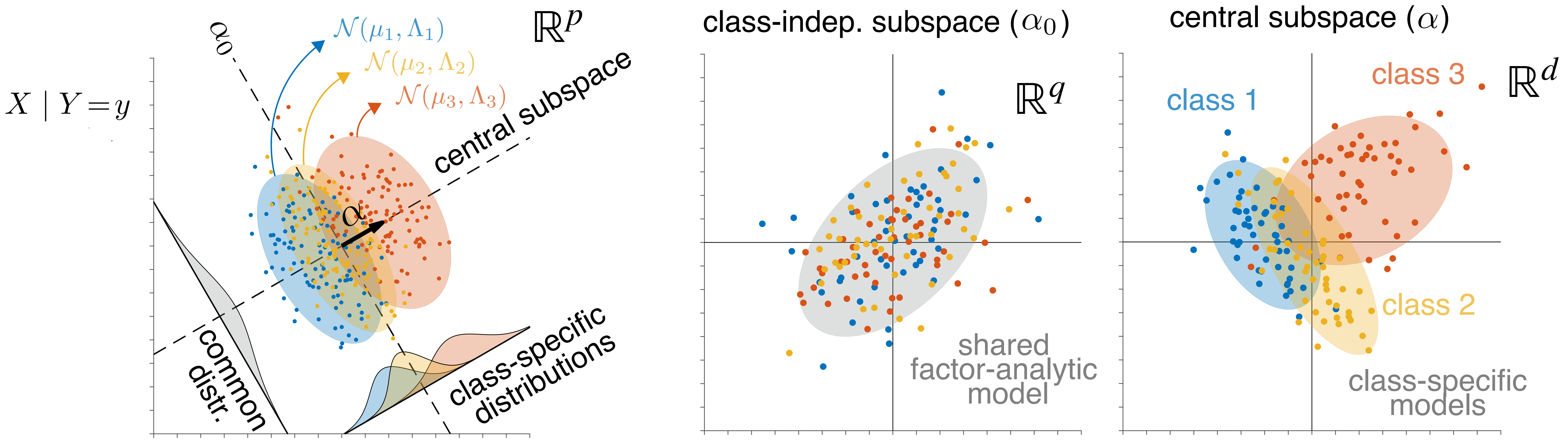


$$X \mid Y \sim \mathcal{N}(\alpha\nu_y, \alpha\Sigma_y\alpha^\top + \alpha_0\Sigma_0\alpha_0^\top)$$

Class-specific covariance

Shared covariance

Class-Conditional Factor Analytic Dimensions (CFAD)



$$X \mid Y \sim \mathcal{N}(\alpha \nu_y, \alpha \Sigma_y \alpha^\top + \alpha_0 \Sigma_0 \alpha_0^\top + \Psi)$$

Class-specific covariance | Shared covariance | Independent Noise

Inference

CFAD: $X \mid Y \sim \mathcal{N}(\alpha\nu_y, \alpha\Sigma_y\alpha^\top + \alpha_0\Sigma_0\alpha_0^\top + \Psi)$

Maximize: $\mathscr{L}(\alpha, \alpha_0, \Sigma_y, \Sigma_0, \Psi)$

Orthonormal matrices: α, α_0

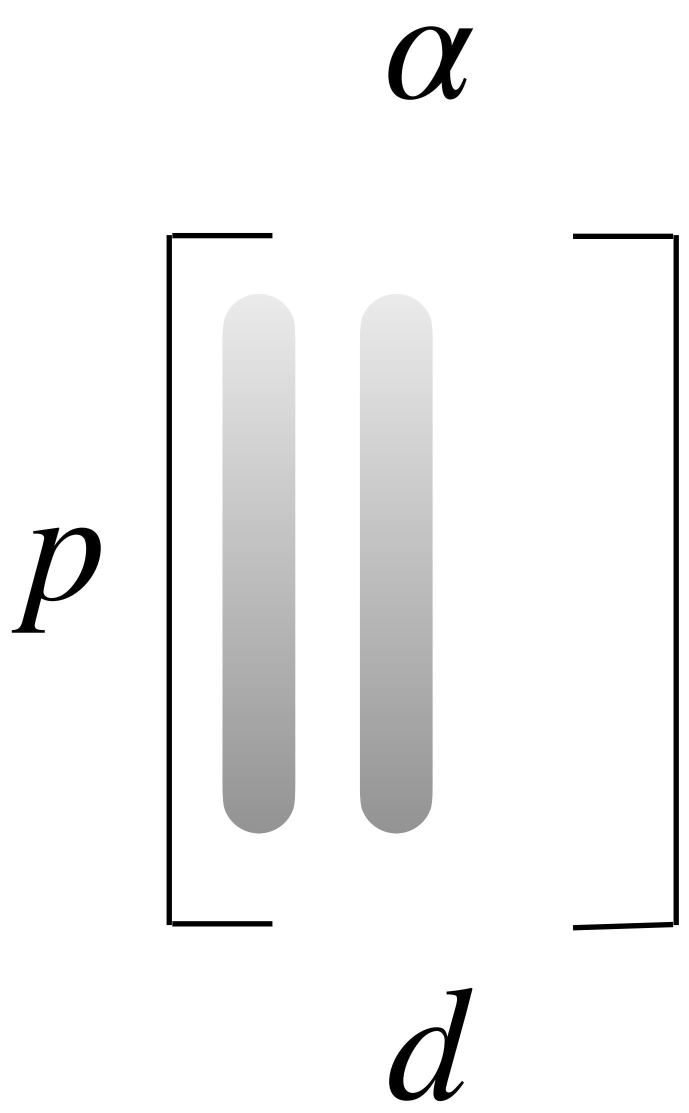
$\alpha \perp \alpha_0$

Riemannian Optimization

smooth-CFAD

Maximize: $\mathscr{L}(\alpha, \alpha_0, \Sigma_y, \Sigma_0, \Psi)$ + log(prior)

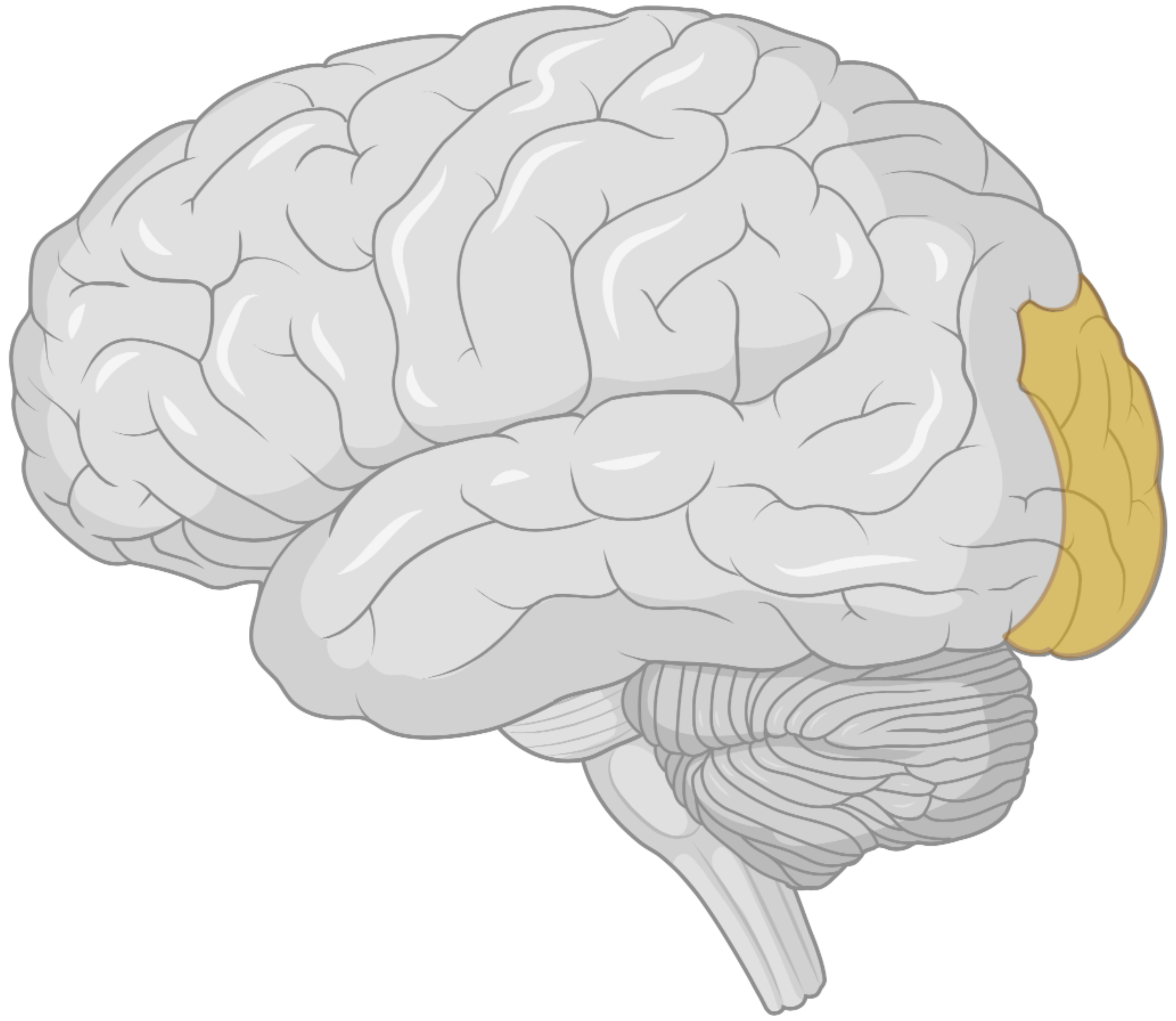
smooth-CFAD



Maximize: $\mathscr{L}(\alpha, \alpha_0, \Sigma_y, \Sigma_0, \Psi) + \lambda \frac{1}{2} \mathrm{Tr}(\alpha^\top D \alpha)$

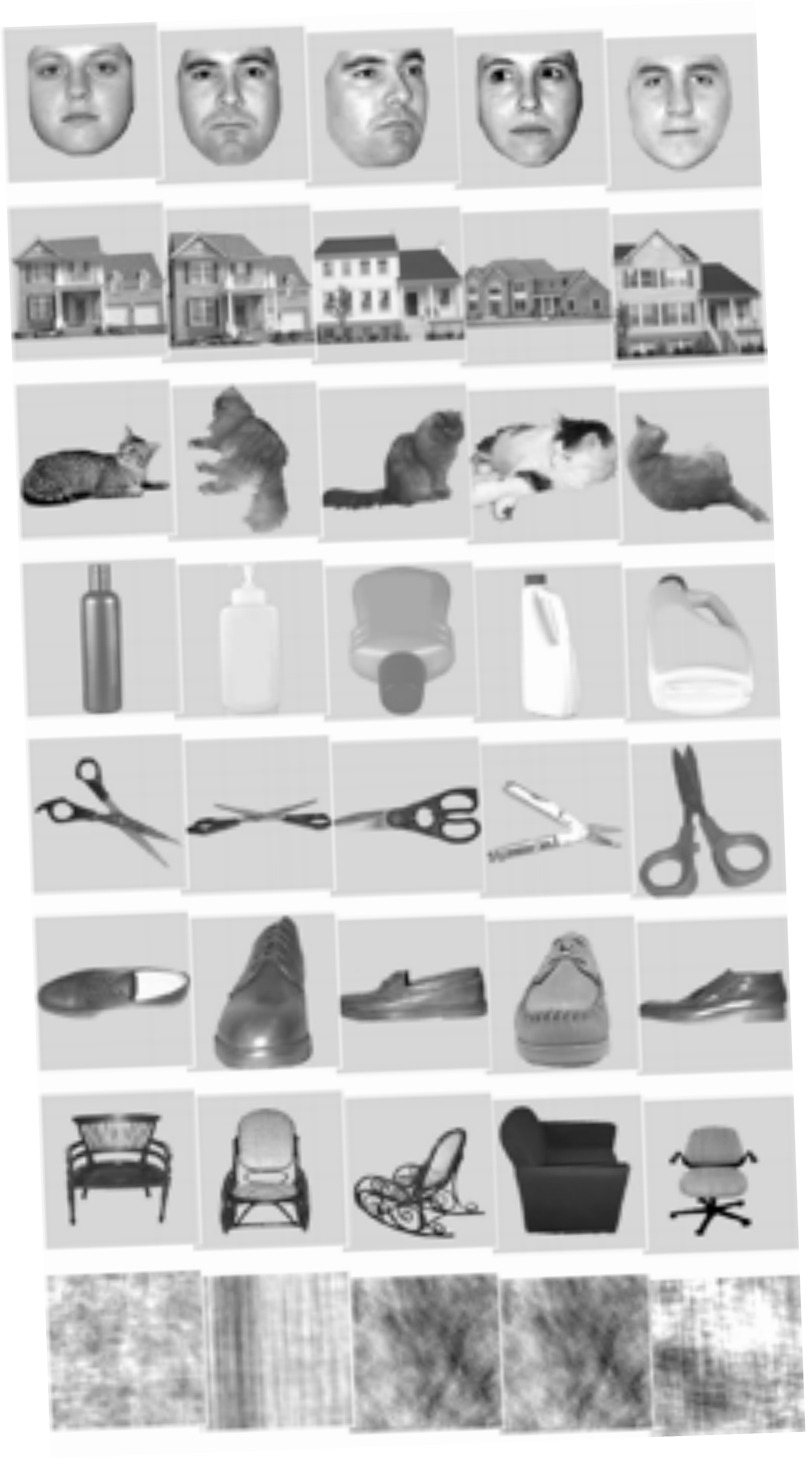
Smoothness prior

Visual Object Recognition



$$X \to \alpha^\top X \to Y$$

$\mathbb{R}^p$ $\mathbb{R}^d$



Haxby et. al., 2001

Visual Object Recognition

$N = 100$ per class, $p \in (307 - 675)$

8-class classification accuracy; 12.5% is chance performance

SUBJECT	d	sCFAD	CFAD	LDA	SIR	SAVE	DR	LAD	PCA	LOL[1]	RRR
1	10	**62.4**	57.3	59.3	59.5	10.0	54.1	52.1	21.9	30.1	23.0
2	10	**71.8**	68.9	58.9	59.9	12.1	62.3	36.5	23.5	31.3	18.7
3	10	**66.4**	63.0	60.3	62.1	10.2	61.7	44.0	32.8	42.3	16.0
4	20	**62.2**	61.2	22.0	20.8	11.6	30.3	29.3	24.8	26.8	19.6
5	10	**72.8**	69.8	60.1	61.8	12.2	63.5	50.8	34.7	41.2	18.1
6	10	**73.1**	70.9	71.5	70.8	10.7	71.9	65.0	39.7	53.0	21.2

Existing SDR methods[2] (SIR, SAVE, DR, LAD)

1. Vogelstein et. al., 2017
2. Cook, 2007; Cook & Forzani, 2009

HCP Working Memory Task

$N = 132$ per class, $p = 3093 \qquad (N < p)$

4-class classification accuracy; 12.5% is chance performance

Subject	sCFAD		CFAD	LDA	PCA	LOL	RRR
	d	%	%	%	%	%	%
1	10	**73.9**	70.9	64.7	67.4	72.7	20.3
2	10	**85.5**	83.3	84.8	77.3	82.7	24.4
3	10	**93.2**	91.8	88.6	80.1	85.0	20.8
4	20	**86.2**	86.0	82.2	82.6	84.8	24.5
5	10	85.1	83.5	**86.3**	82.2	82.0	25.2
6	10	**94.1**	93.0	87.1	85.2	87.5	25.0
7	10	**91.2**	89.9	88.4	88.2	91.0	26.7
8	10	**87.1**	83.3	81.6	82.6	82.7	25.2
9	10	**89.9**	87.3	87.8	79.2	79.9	22.5
10	10	**92.8**	89.8	92.6	86.6	87.9	29.2

Barch et. al., 2013

Conclusion

CFAD adds to the limited literature on high-dimensional small-sample size data.

It shows improved classification performance on real-world fMRI datasets.