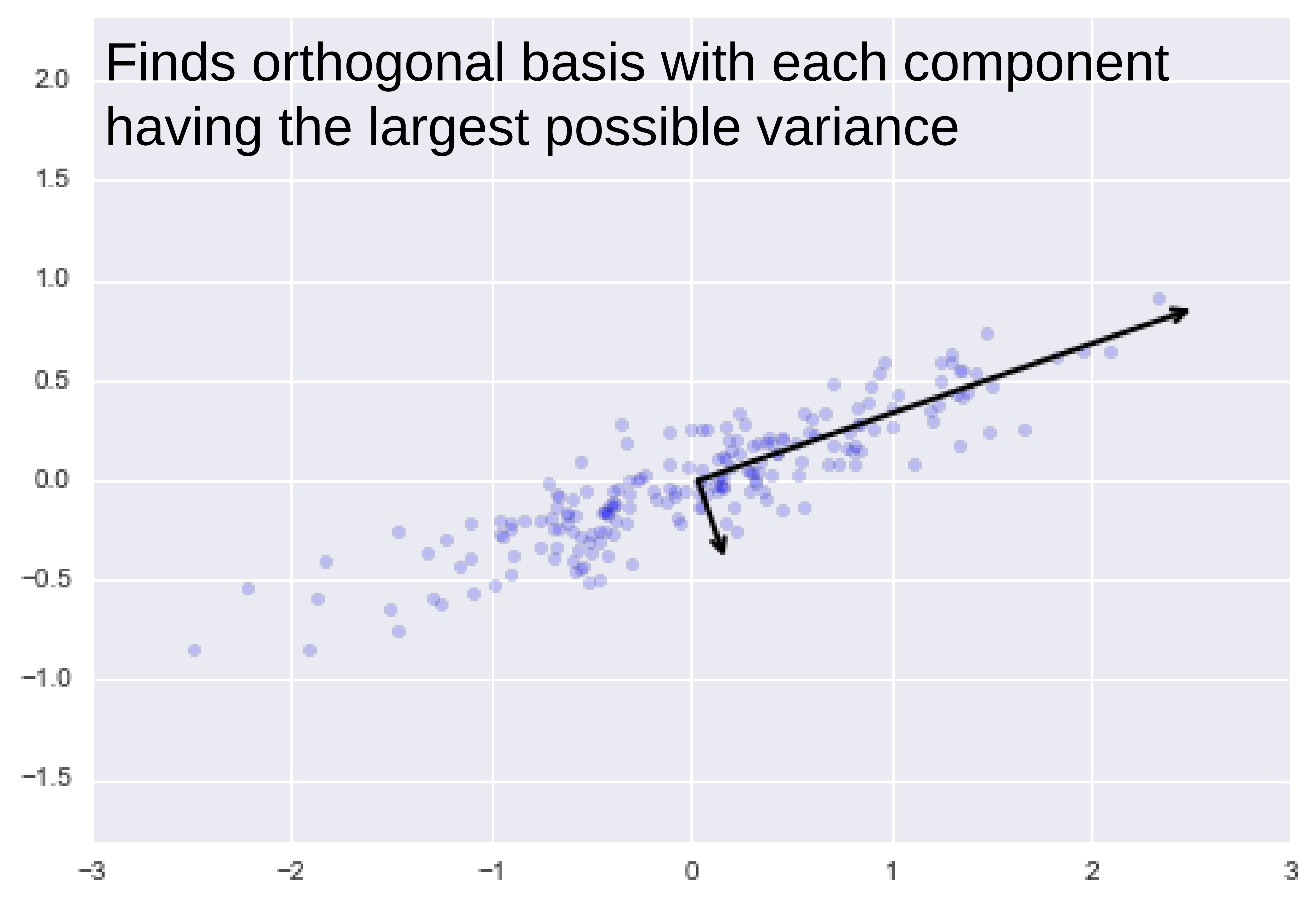
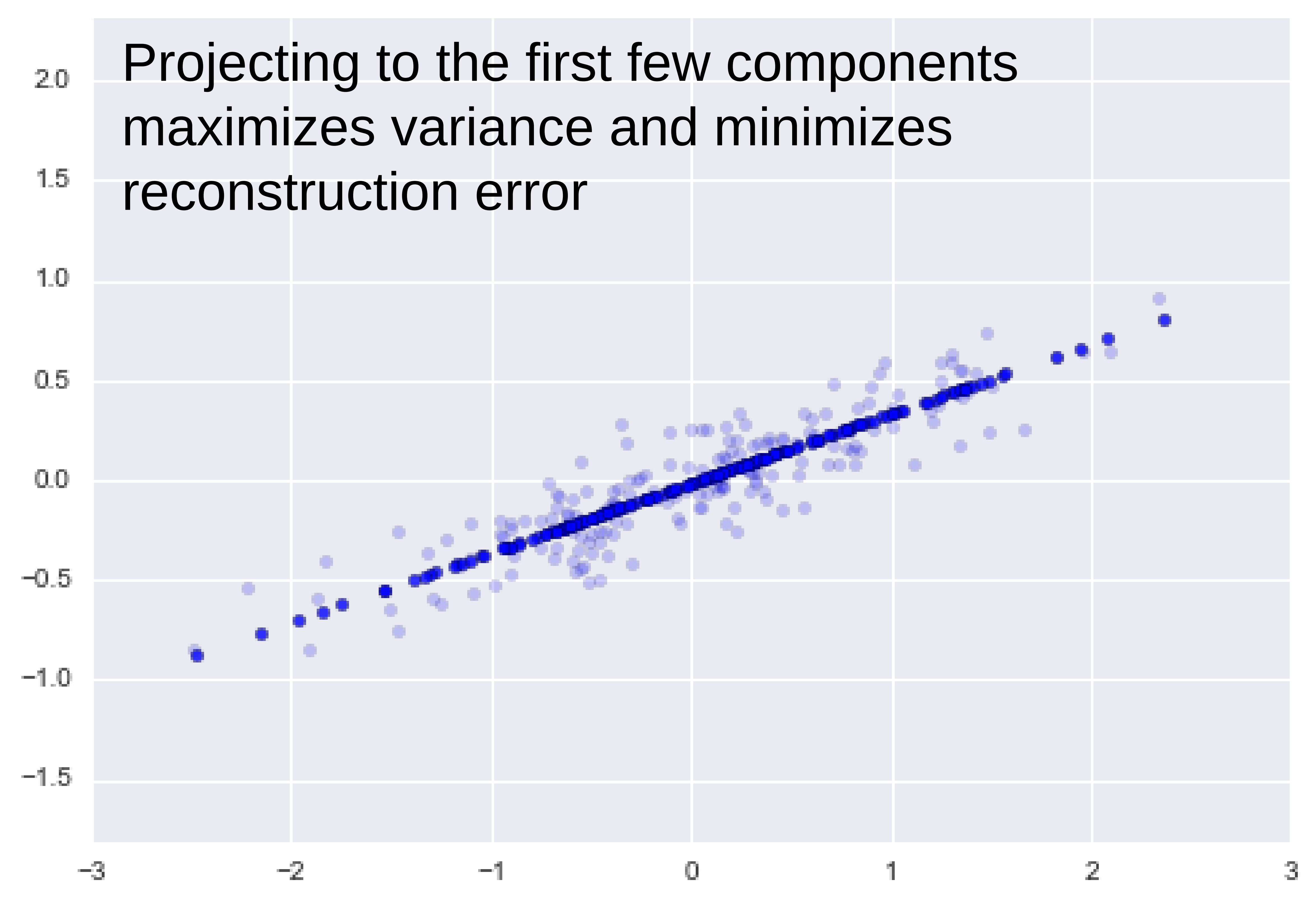
Dimensionality Reduction of SDSS Spectra with Autoencoders

Stephen Portillo, DIRAC Institute, UW
with Andy Connolly
Astroinformatics 2019, 25 June 2019

Principal Component Analysis

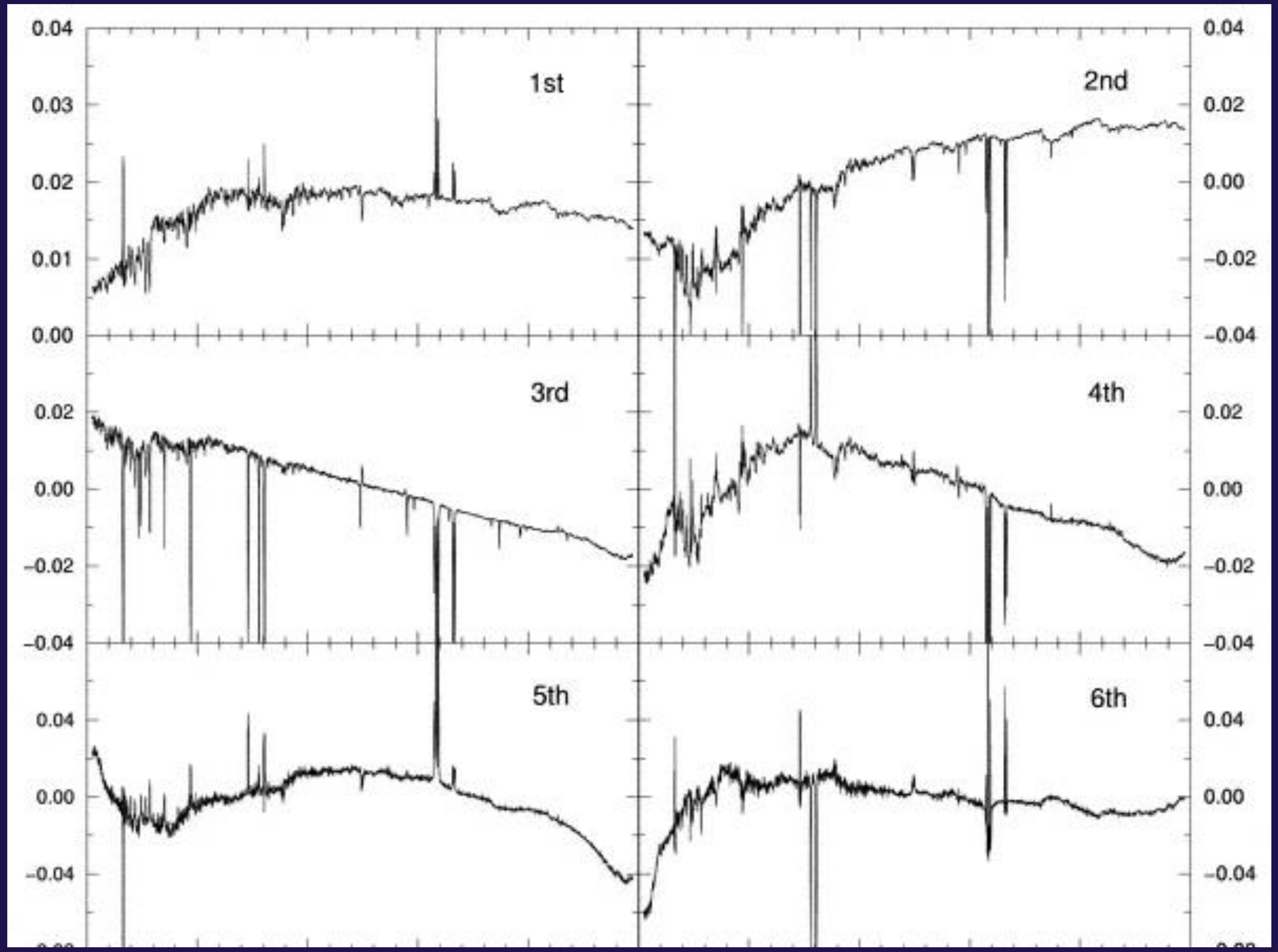




Jake VanderPlas, Python Data Science Handbook

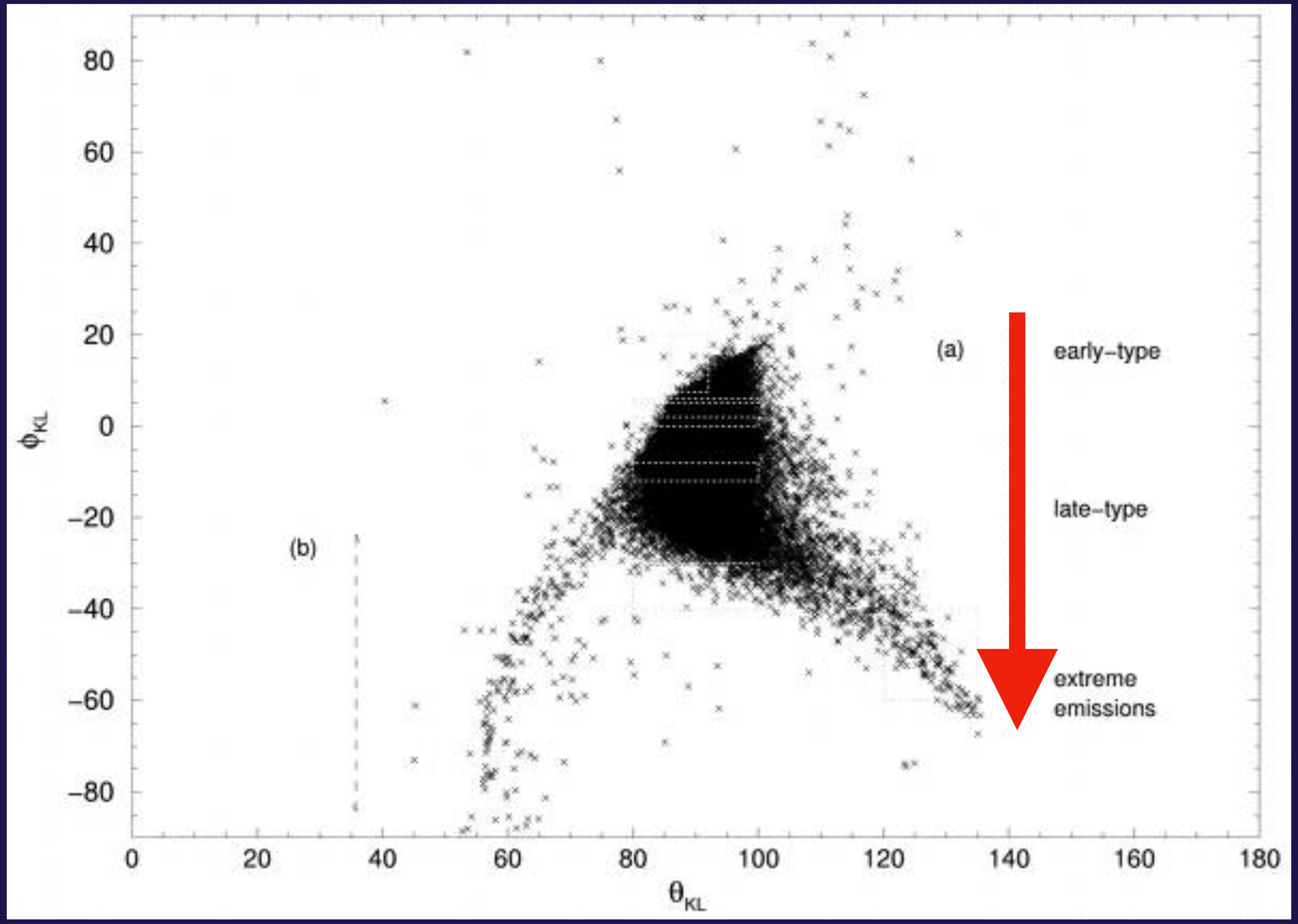
Principal Component Analysis

Eigenspectra can be interpreted



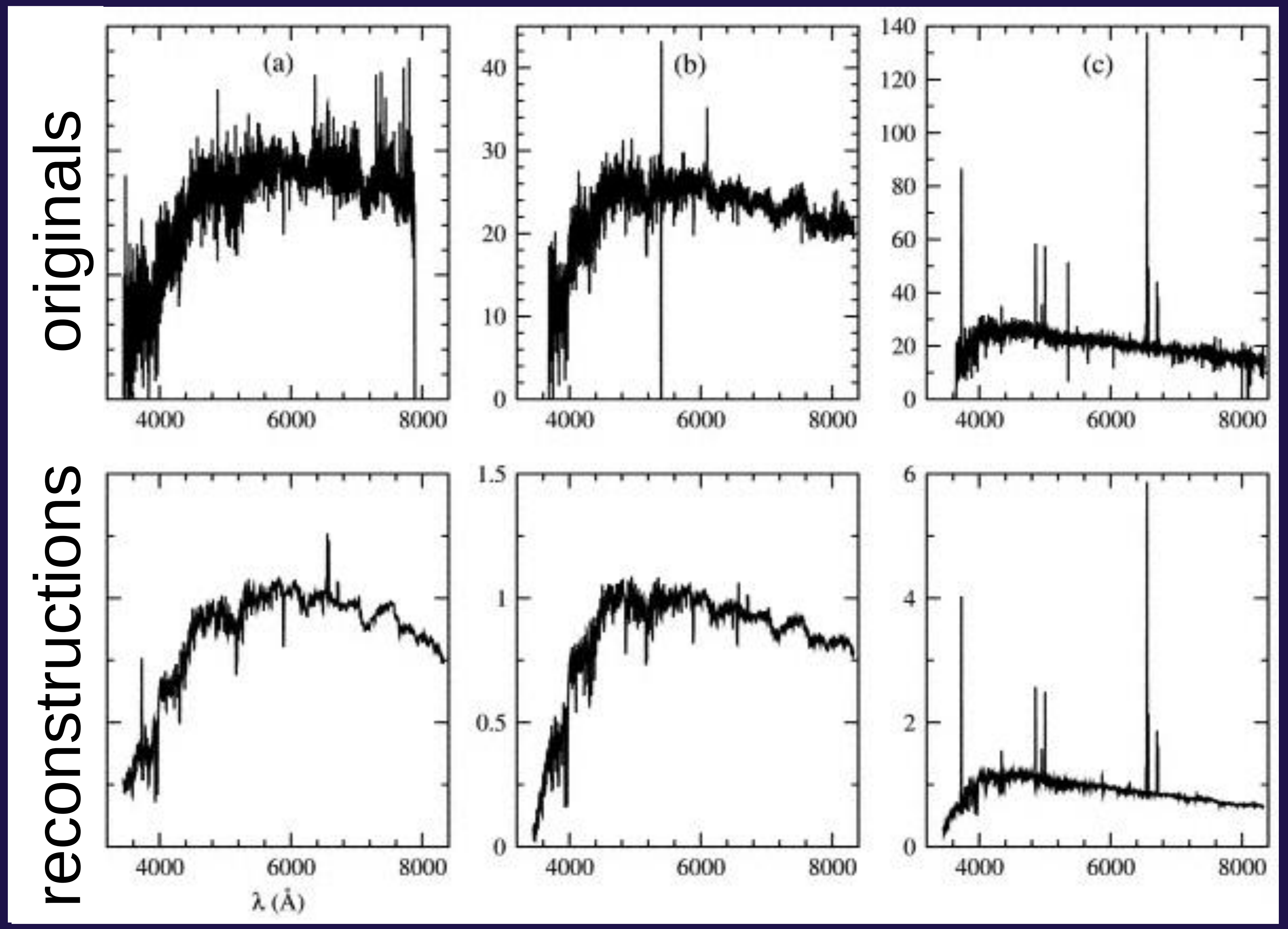
Yip et al. (2004)

First 3 coefficients can be used to separate classes of galaxies



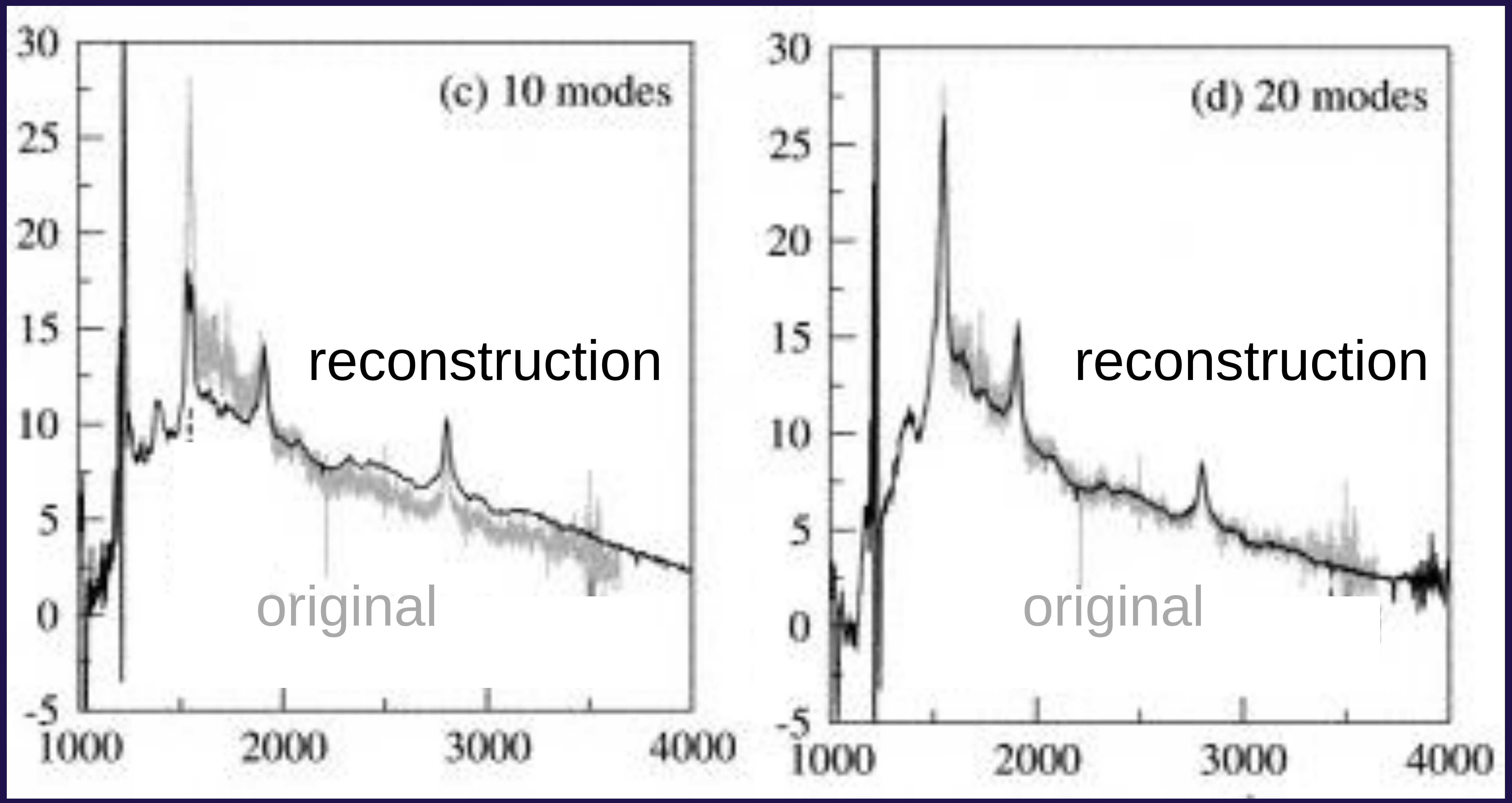
Principal Component Analysis

Projection can de-noise spectra



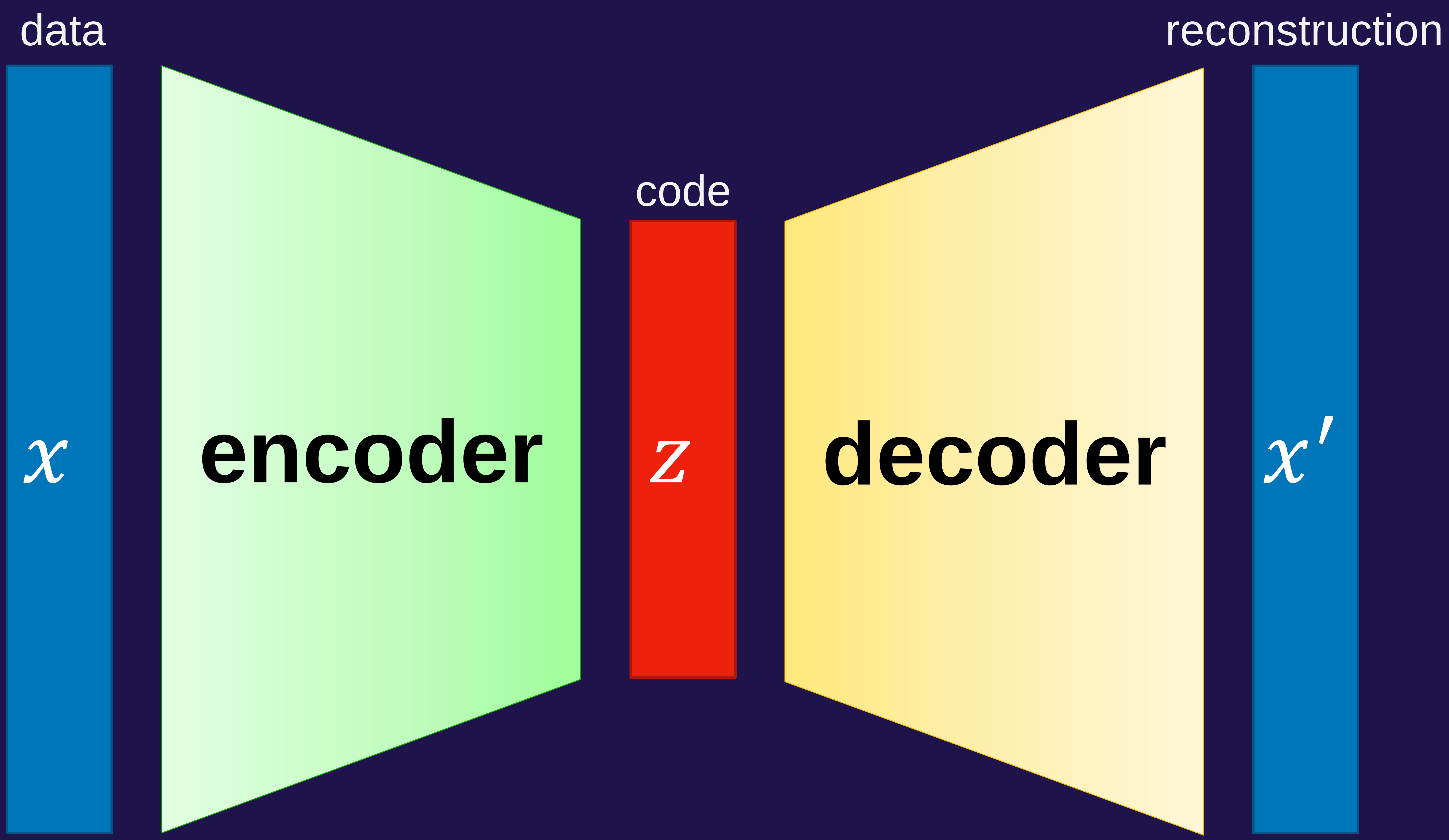
Yip et al. (2004)

Non-linear features (eg. broad lines) can take many components to reconstruct



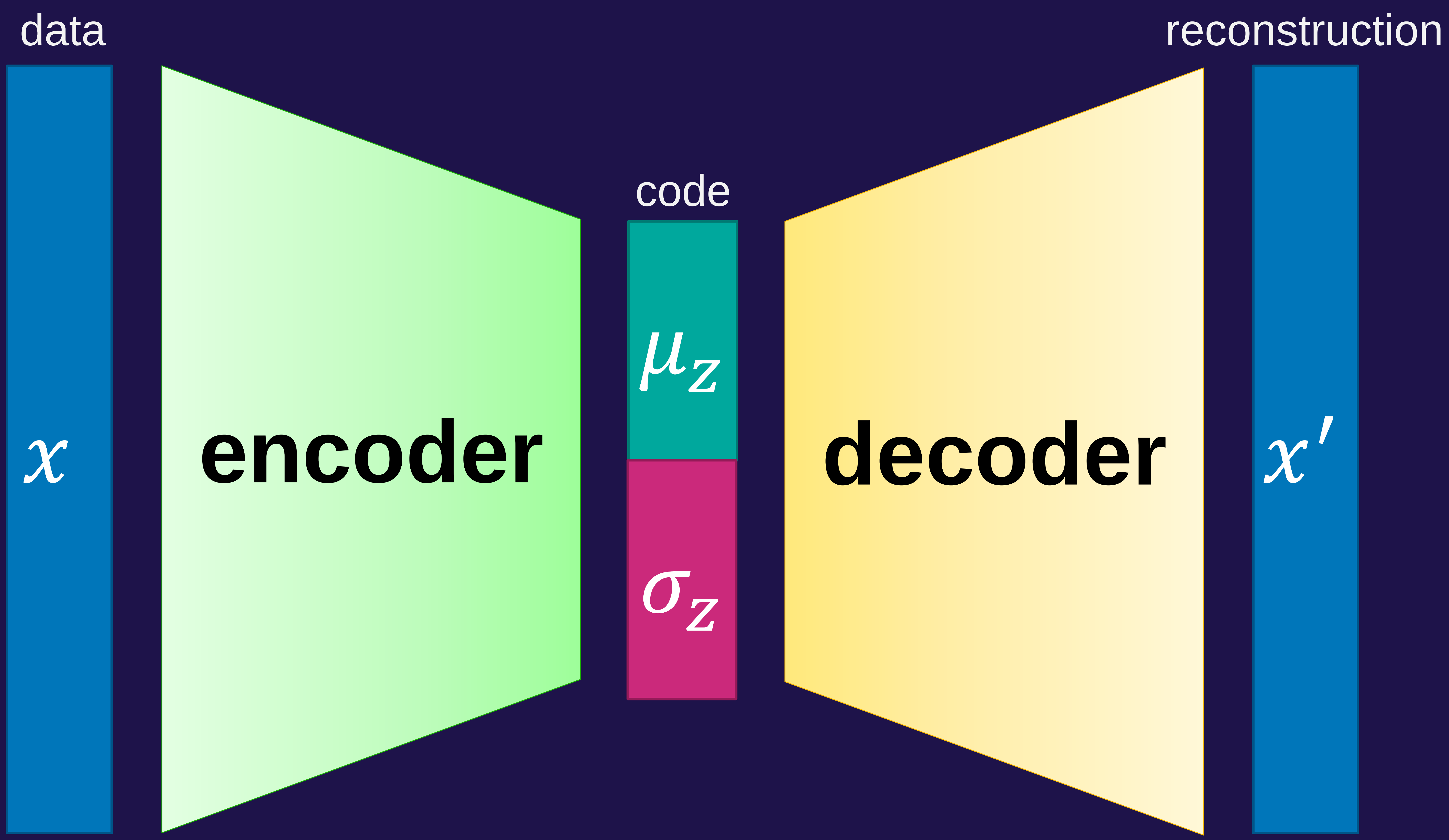
Autoencoders

- Trained to minimize reconstruction loss $\mathcal{L}(x, x')$
- Code z is a compression that is tailored to the data
- A non-linear generalization of PCA



Variational Autoencoders

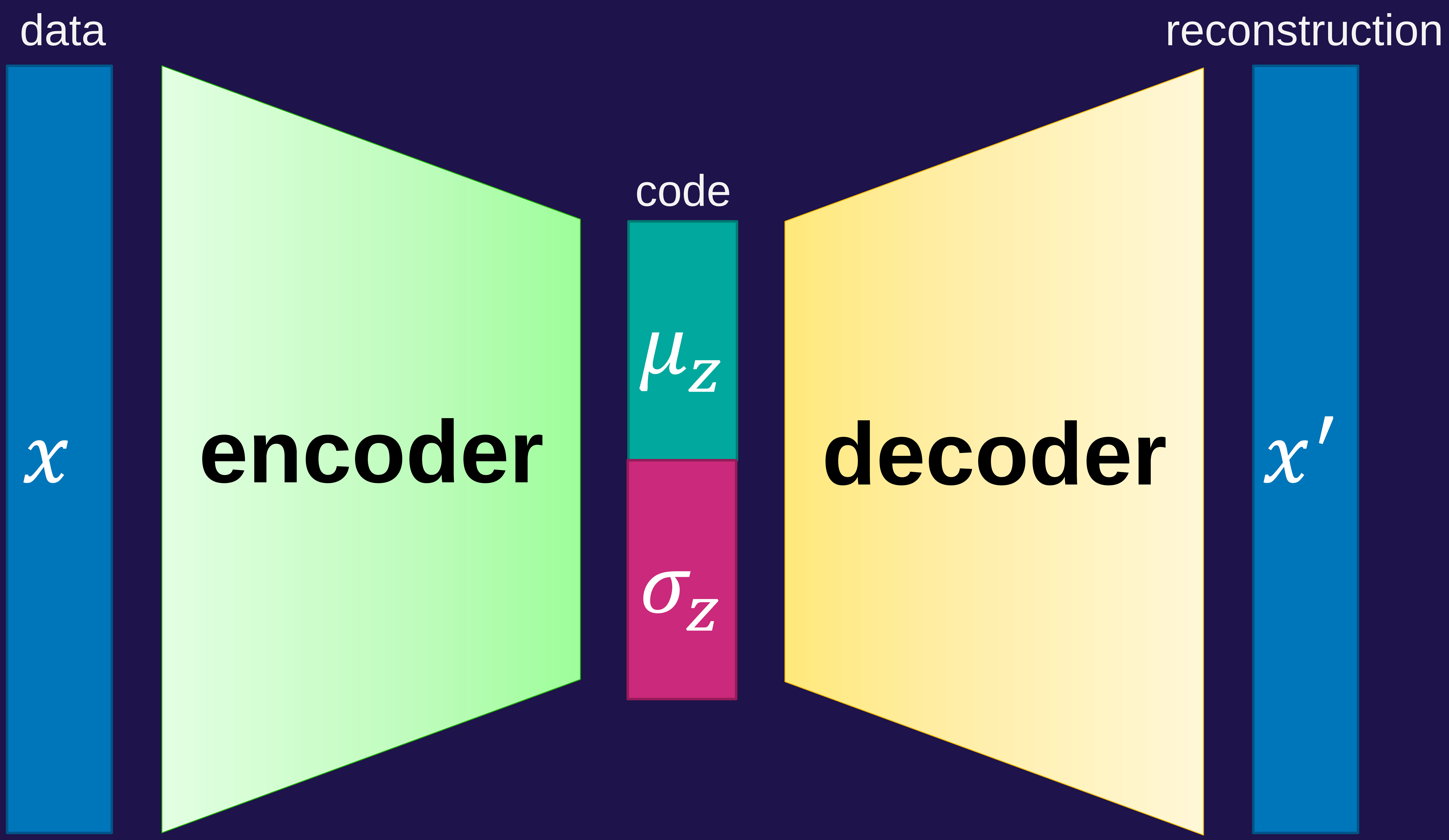
- Decoder is a generative model with parameters z
- Reconstruction loss is a likelihood
  $P(x|z) = \mathcal{L}(x, x')$
- Placing a prior on z implies a posterior $P(z|x)$
- Encoder approximates posterior with $\mathcal{N}(\mu_z, \sigma_z)$ - variational inference!



Kingma & Welling, ICML 2014

Application to SDSS

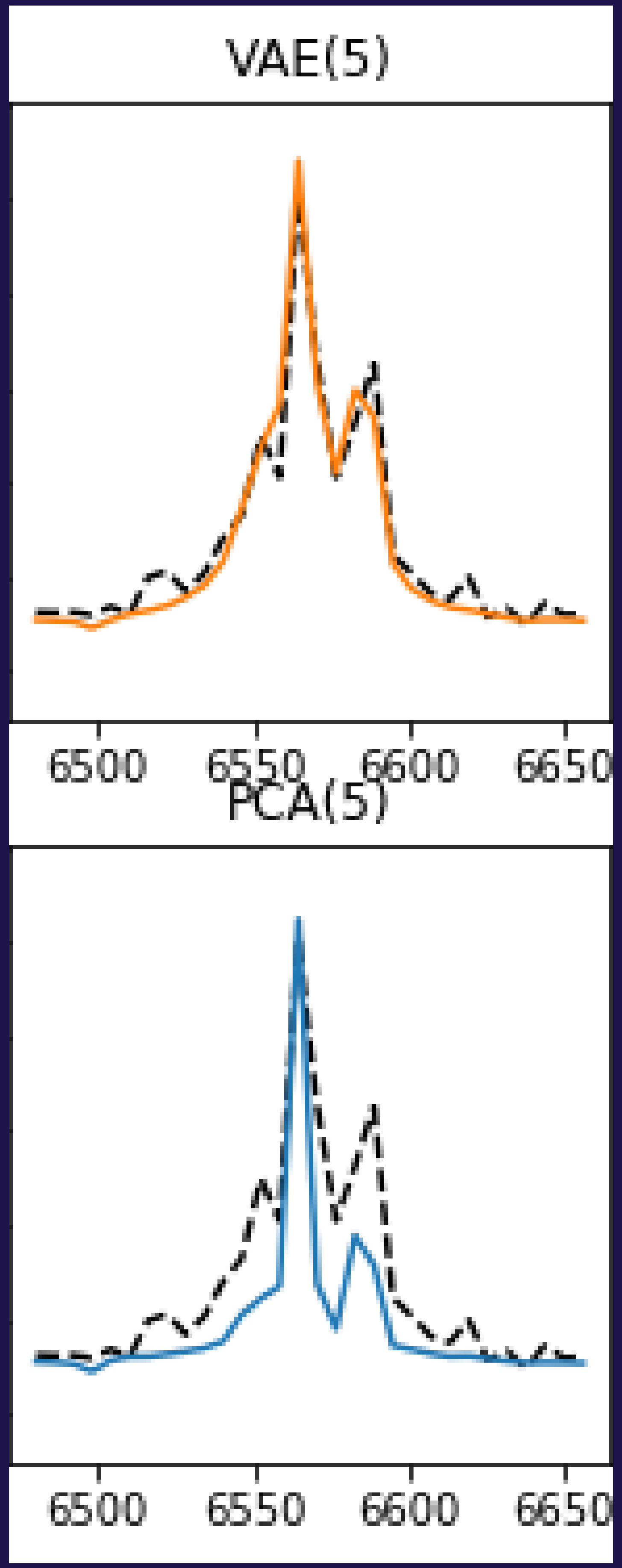
- Spectra de-redshifted to rest frame, rebinned to 1K
- Encoder and decoder are two layers (167, 61) with nodes
- Latent space of 10 parameters
- Trained on 23K spectra



Reconstruction

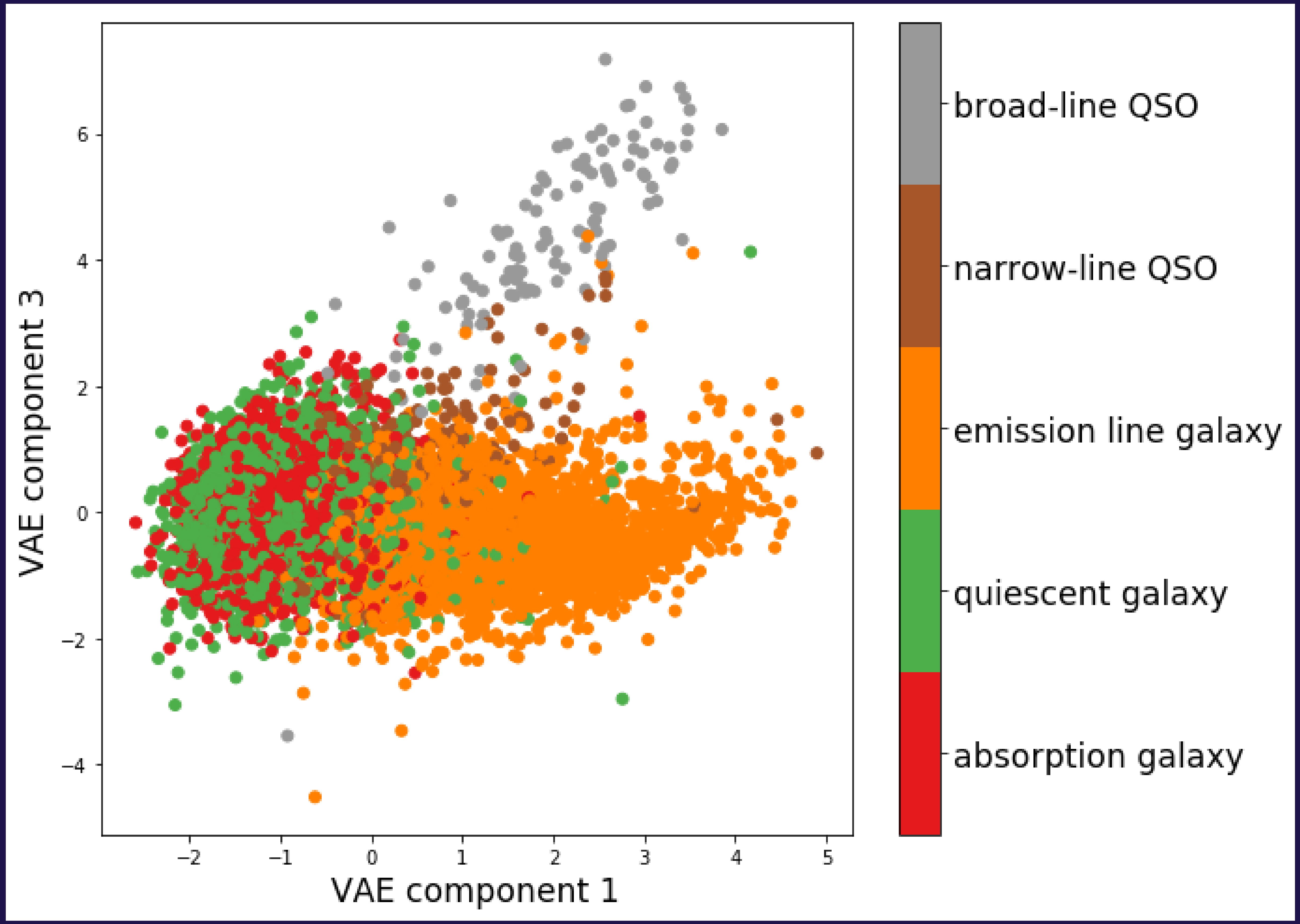
VAE with 10 components outperforms PCA with 10 components

Class	MSE Improvement
All	2.6%
Absorption galaxy	3.3%
Quiescent galaxy	1.6%
Emission line galaxy	0.8%
Narrow-line QSO	1.5%
Broad-line QSO	23.8%



VAE is able to reconstruct broad spectral lines with fewer components

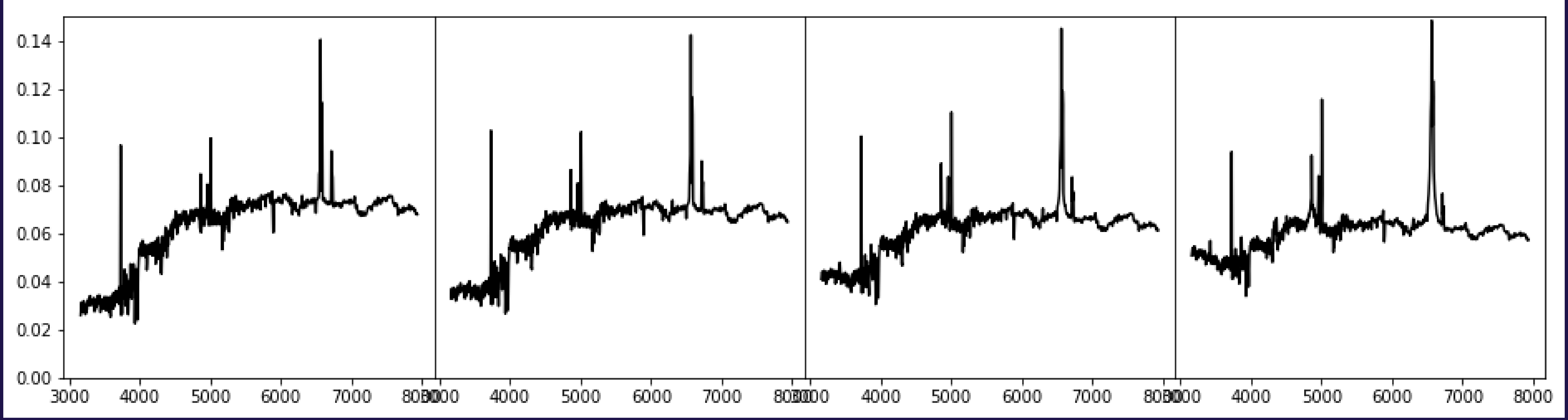
Separation of Classes



First 3 VAE components separate classes well

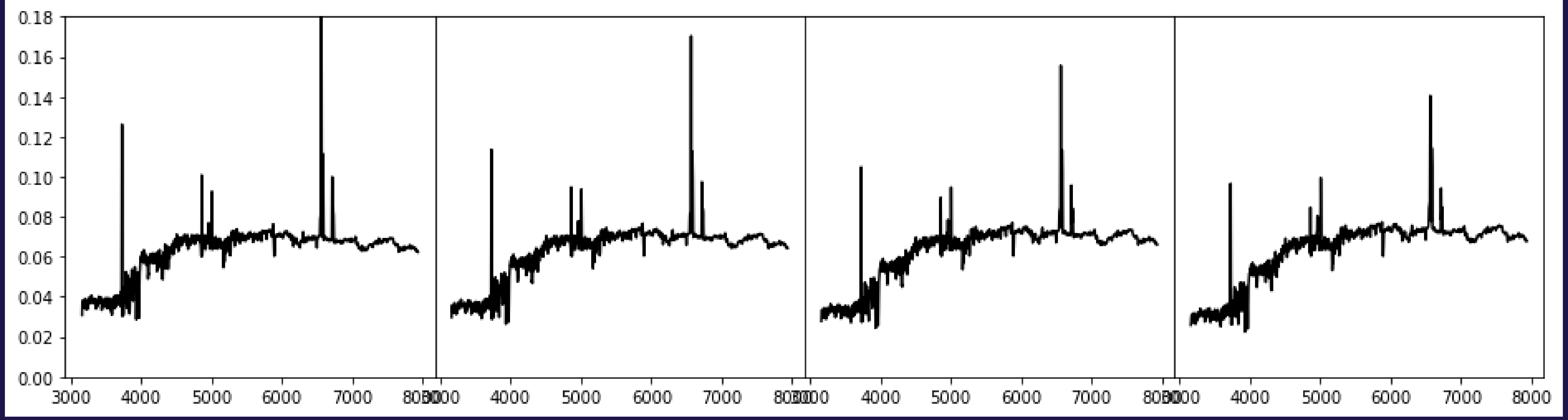
Narrow-line → Broad-line QSO

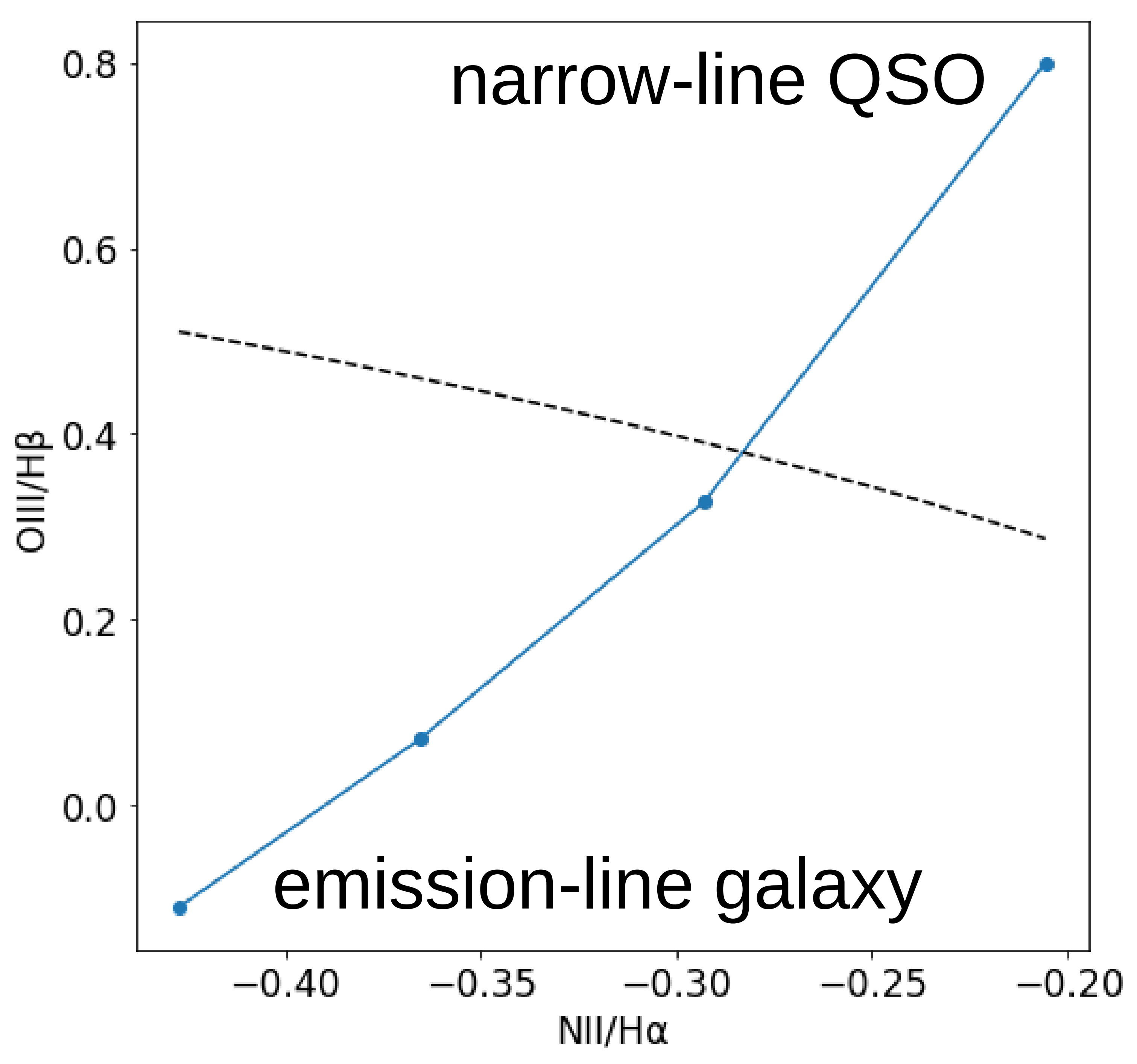
Lines broaden, blue continuum rises, line ratios change



Emission Line Galaxy → Narrow-line QSO

OIII/Hβ and NII/Hα ratios change, in agreement with Kewley et al. (2001)



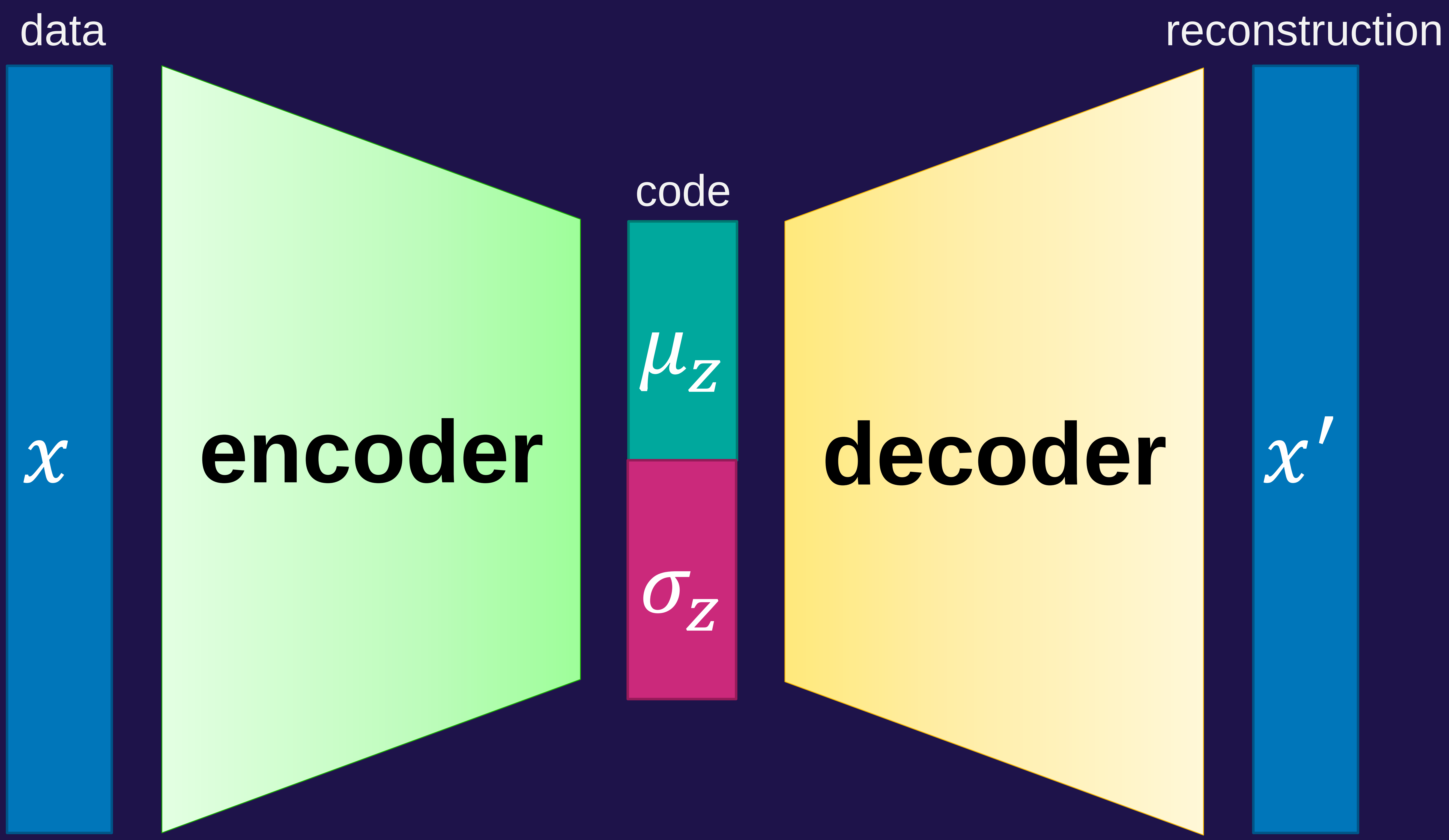


Conclusion

- Autoencoders are an unsupervised method to learn compressions of data sets, and can be thought of as a non-linear generalization of PCA
- Autoencoders have better reconstruction performance than PCA, especially for non-linear features like broad spectral lines
- The autoencoder latent space separates galaxy classes
- Traversing the latent space generates series of synthetic spectra that change in physically plausible ways

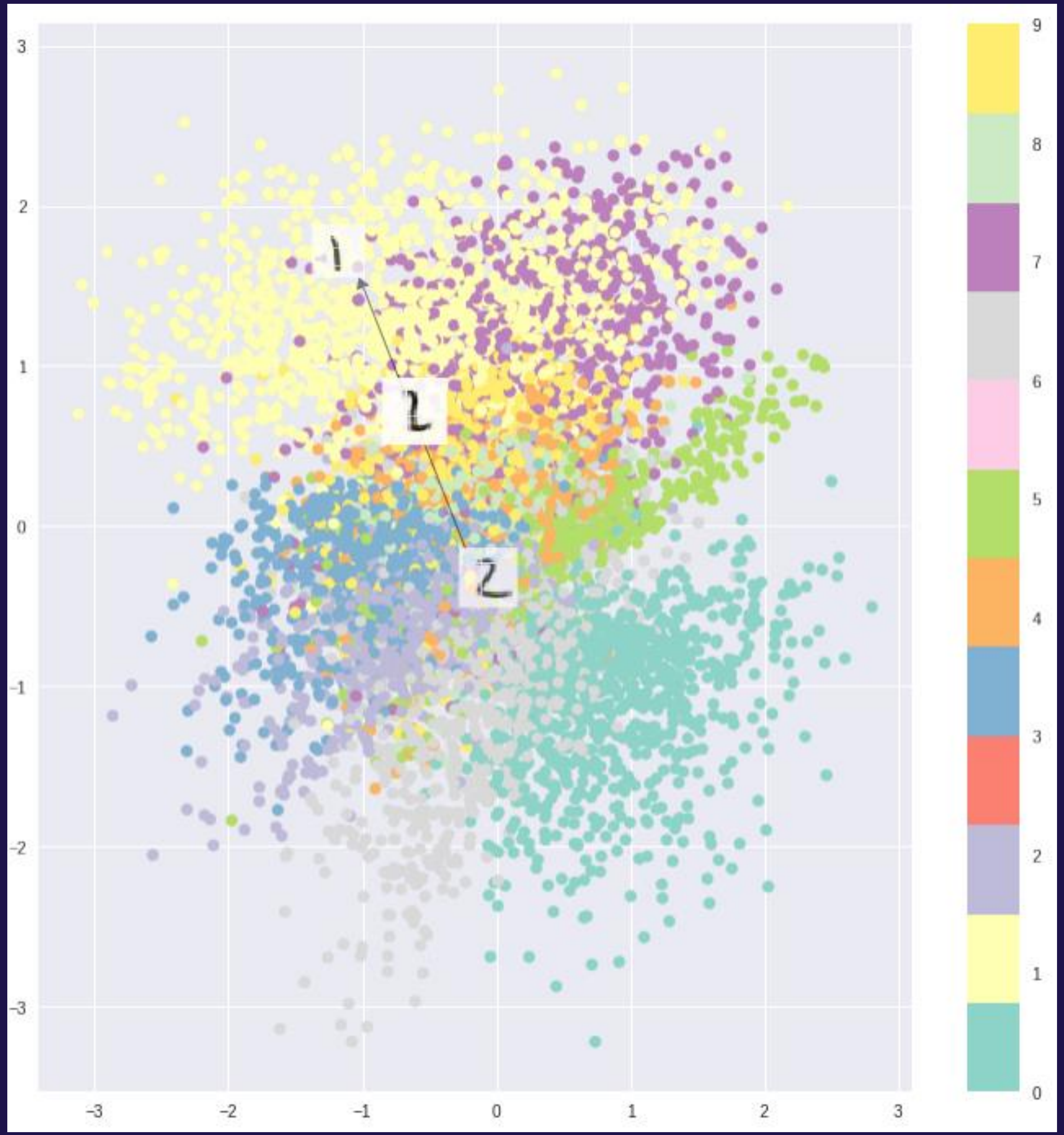
Variational Autoencoders

- Results in an additional loss term that depends on encoder and prior π
  $KL\big(\mathcal{N}(\mu_z|\sigma_z)||\pi(z)\big)$
- InfoVAE changes loss function to prefer mutual information between data and code



Kingma & Welling, ICML 2014
Zhao et al., arXiv:1706.02262

Variational Autoencoders



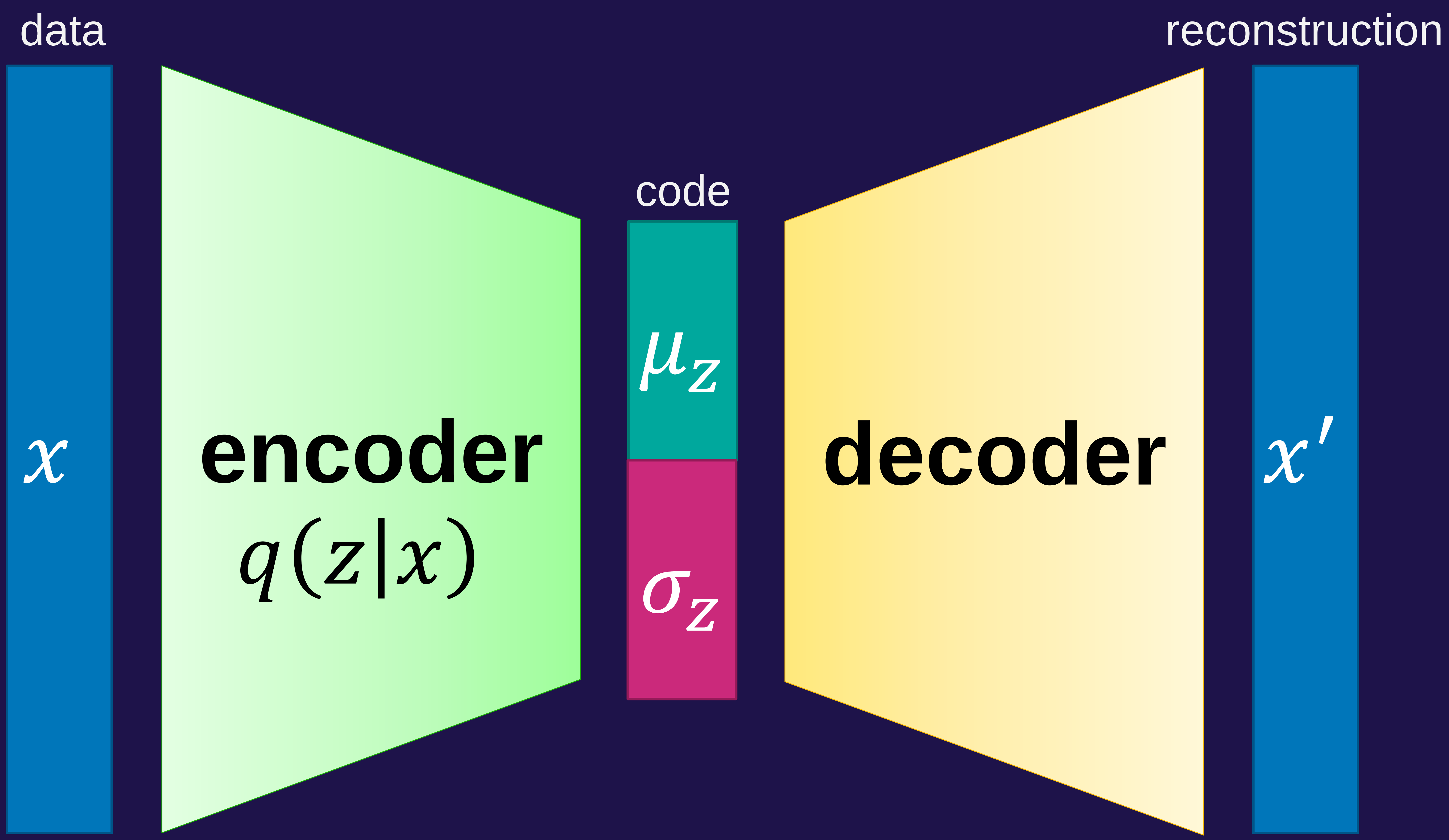
Penalizing KL divergence between encoder and prior encourages encodings to stay together

Irhum Shafkat, “Intuitively Understanding Variational Autoencoders”

InfoVAE Loss

- InfoVAE changes loss function to prefer mutual information between data and code

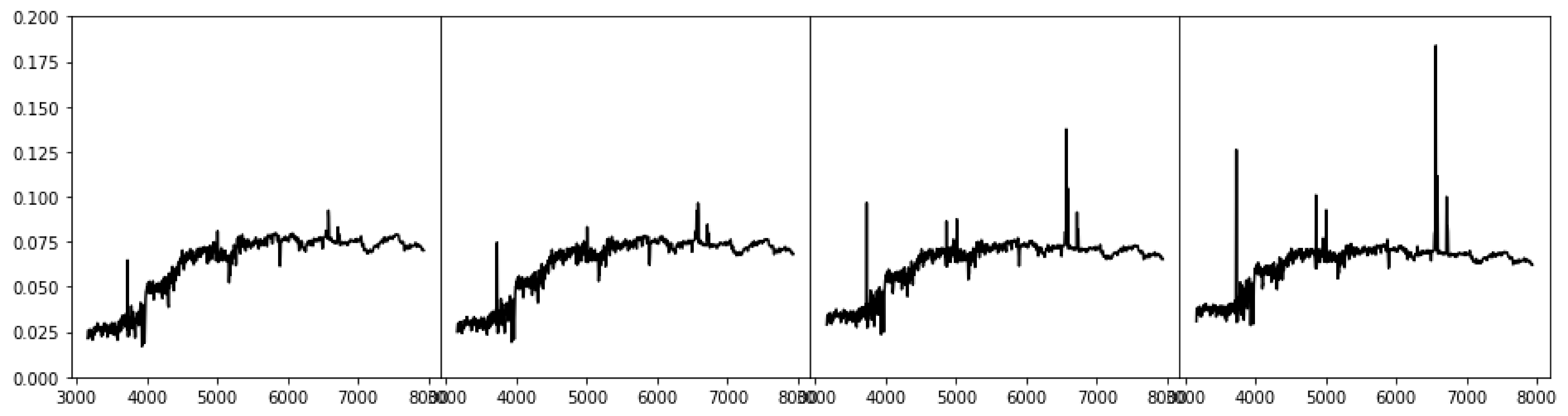
$$MMD\big(q(z)||\pi(z)\big)$$



Zhao et al., arXiv:1706.02262

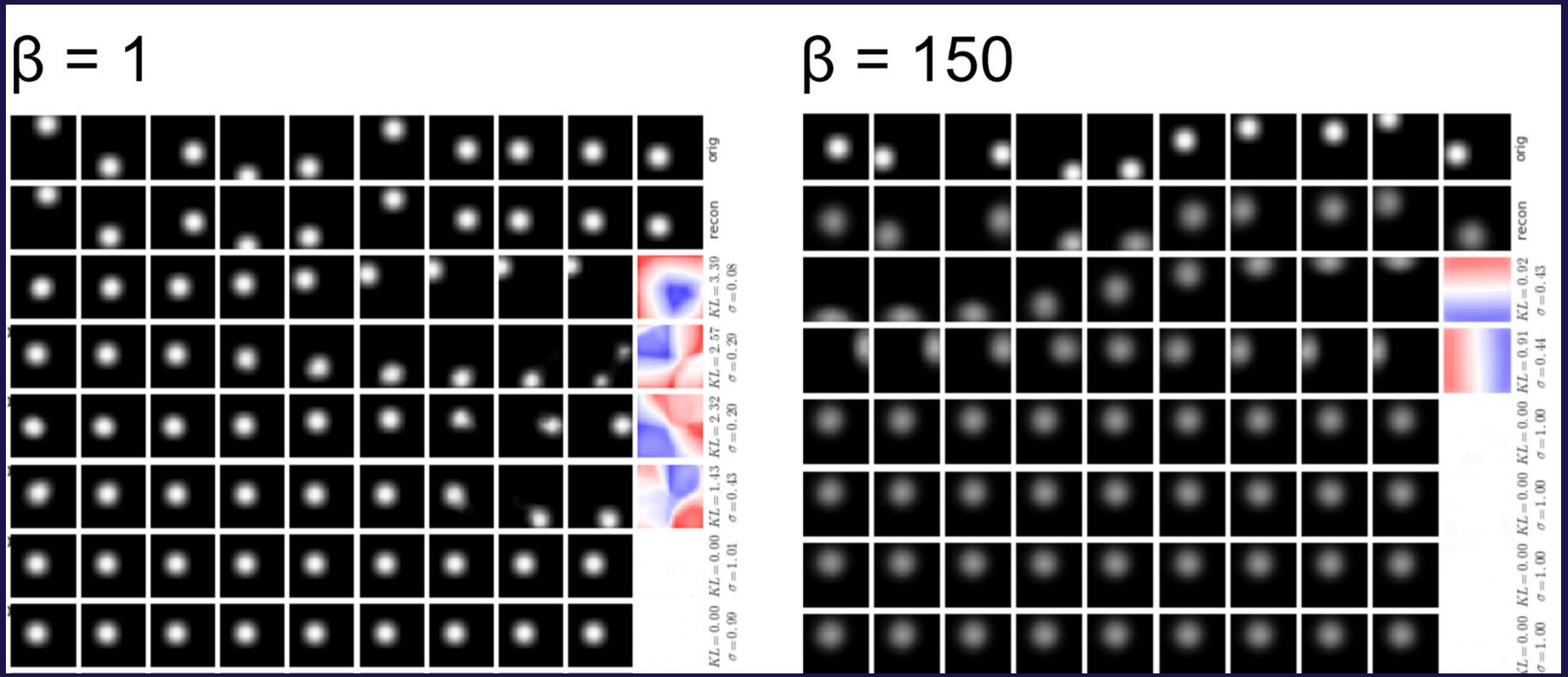
Absorption → Emission Line Galaxy

Emission lines and blue continuum rise



Disentangled Variational Autoencoder (β-VAE)

- Upweight the KL divergence contribution to the loss function by multiplying it by $\beta > 1$
- Encourages the encoder to only differ from the prior when it really needs to – using fewer dimensions of latent space



Burgess et al., NIPS 2017