NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas

Department : E & ECE, IIT Kharagpur

Topic

Lecture 57: Variational Autoencoder

CONCEPTS COVERED

Concepts Covered

- ❑ Generative Model
- ❑ Limitations of usual auto-encoder
- ❑ Intuitions behind VAE
- ❑ Variational Inference
- ❑ Practical Realization of VAE

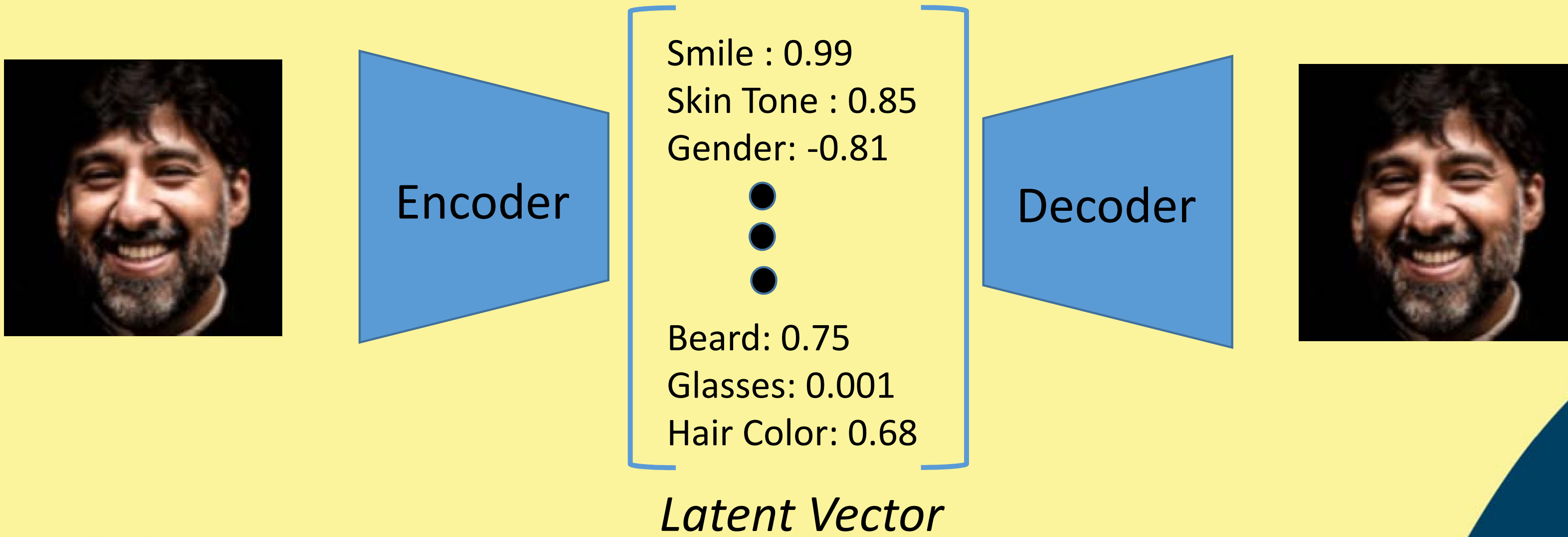
Generative Model

- Big Animal.
- Has four legs.
- Big ears.
- Long trunk.
- A pair of tusks
- …..

Latent Variables

Traditional Autoencoder

- Maps an input image via an encoder to a deterministic latent code
- Decoder maps the latent code to reconstruct the input image

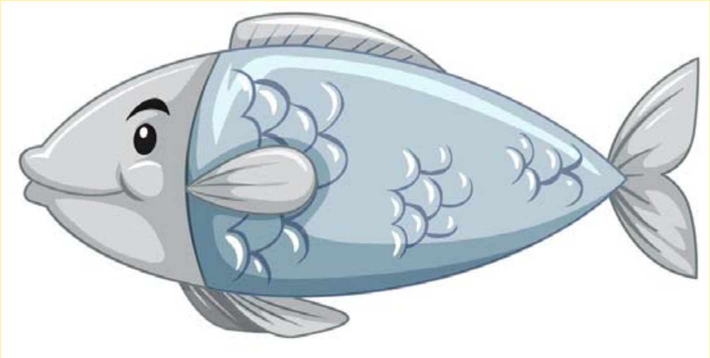


https://www.jeremyjordan.me/variational-autoencoders/

Traditional Autoencoder : Limitations

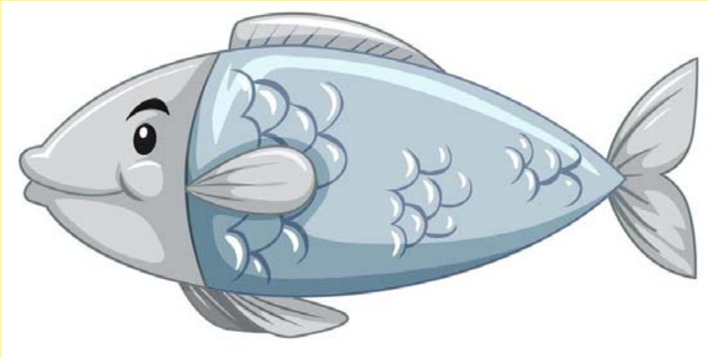
- In pursuit of compact representations, auto-encoders tends to create a latent space which is not continuous
- As a generative model, we need a latent space from which we can smoothly sample and yet get realistic reconstructions
- Auto-encoders do not allow such easy interpolations in latent space

Traditional Autoencoder : Limitations

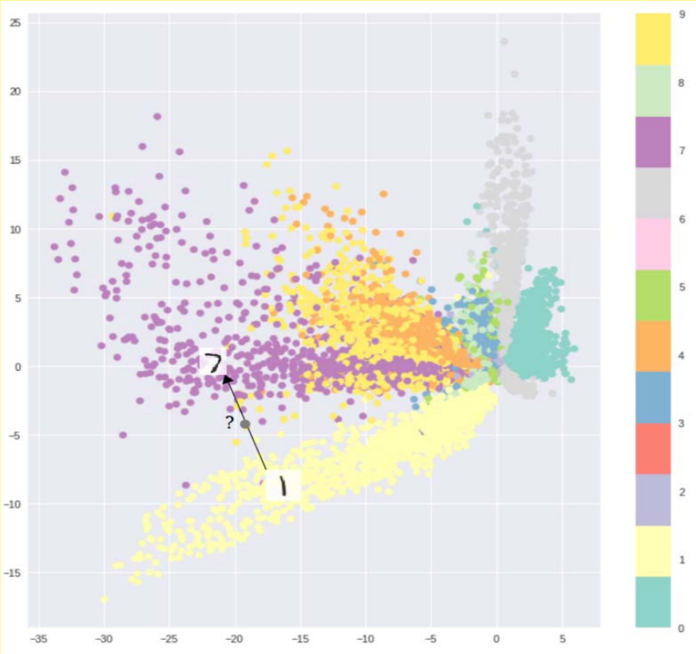


?

Traditional Autoencoder : Limitations

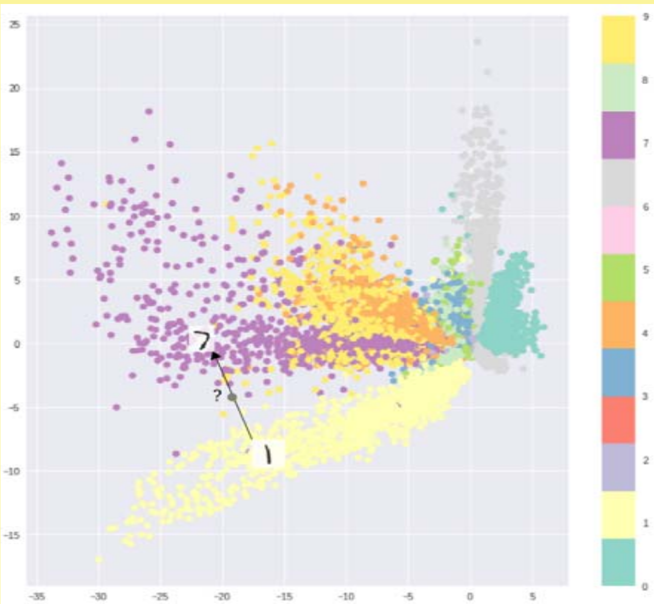


Traditional Autoencoder : Limitations



- Distinct cluster for each class
- Not easy for decoder to reconstruct since we need different distinct codes for each image

Traditional Autoencoder : Limitations



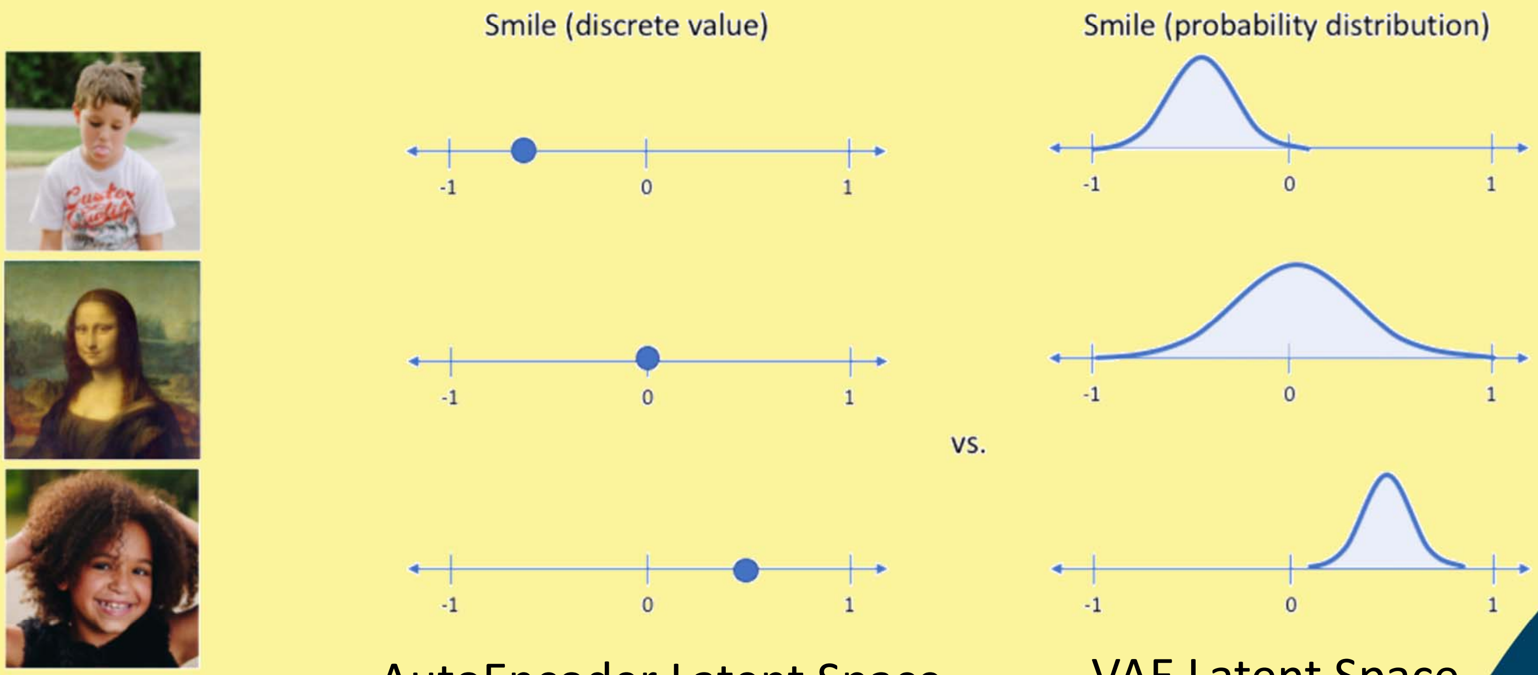
- Discontinuous latent space means decoder never reconstructed from such unexplored points
- If we sample from such points, decoder will give unrealistic output
- **Aim:** Try to make latent space continuous yet maintain the class specific compactness

Image Source: https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

Variational Autoencoder Intuition

- Instead of deterministic latent code we might be interested to learn a distribution over the latent code
- For example, it is more intuitive to determine a range of “smile” value for a face instead of an absolute “smile” value
- Instead of deterministic code, we will now output the mean and standard deviation of each component of the vector (assuming each component is independent of each other)

Autoencoder Intuition vs. VAE Latent Space



AutoEncoder Latent Space

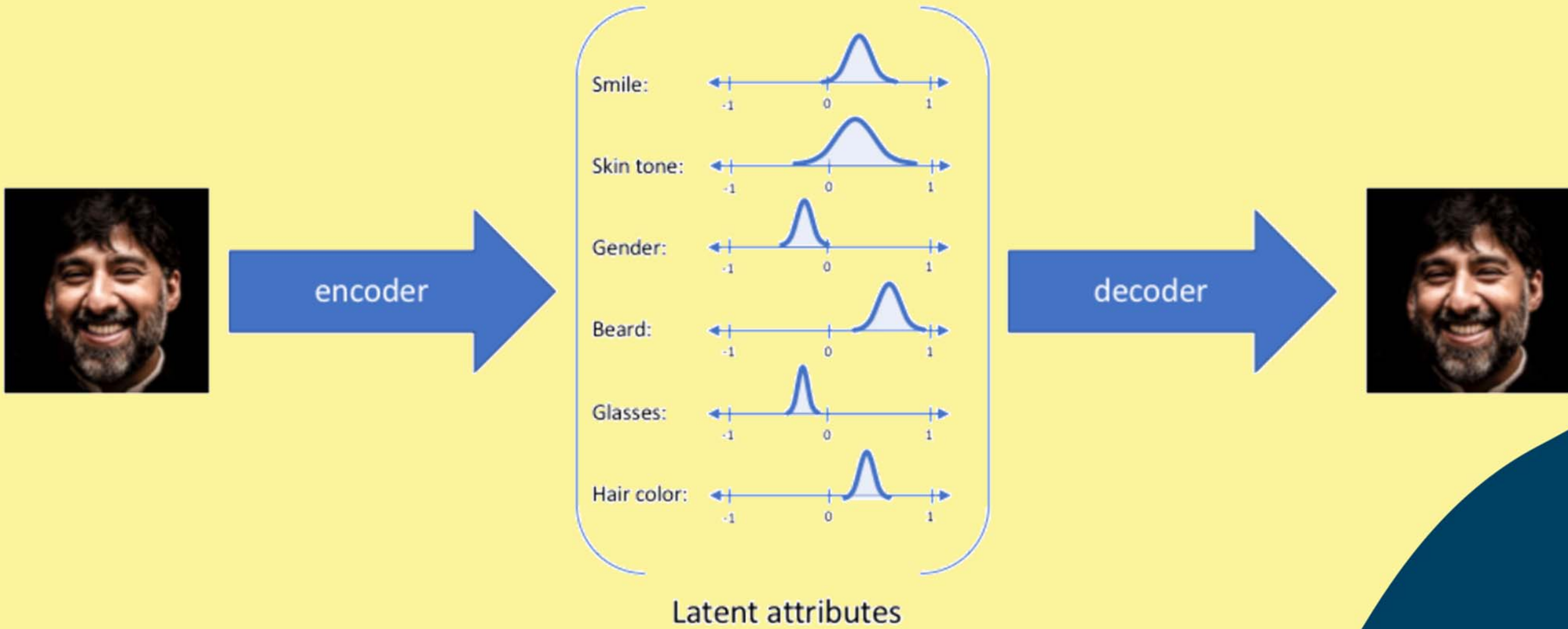
VAE Latent Space

https://www.jeremyjordan.me/variational-autoencoders/

Variational Autoencoder Intuition

- With this setup we can represent each latent factor as a probability distribution
- We can sample from such distribution
- Then the sampled vector can be passed through Decoder (Generator) to generate an image

Variational Autoencoder Intuition



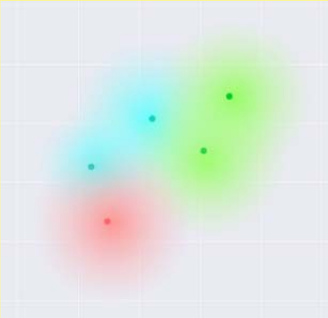
https://www.jeremyjordan.me/variational-autoencoders/

Variational Autoencoder Intuition

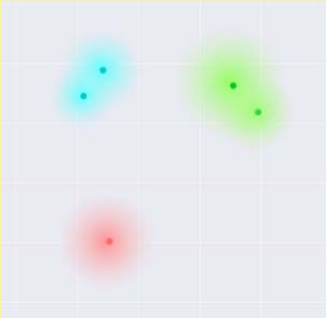
- Mean vector controls where the encoding of an input should be centered around
- Standard deviation controls the "area", how much from the mean the encoding can vary
- As encodings are generated at random from inside a hyper-sphere (distribution) decoder learns that not only is a single point in latent space referring to a sample of that class, but all nearby points refer to the same as well

Variational Autoencoder Intuition

- For smooth interpolations, ideally, we want overlap between samples that are not very similar too, in order to interpolate between classes.
- However μ and σ can take any value and learn to cluster the mean vectors of different classes far apart (and minimize σ) to reduce uncertainty for the Decoder



Our goal



Network might converge to

Image Source: https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

Variational Autoencoder Intuition

- In order to enforce smooth transition we will apply Kullback–Leibler divergence (KL divergence) between the distribution of encoded vectors and a prior distribution asserted on latent distribution space
- KL divergence between two probability distributions simply measures how much they diverge from each other.
- Minimizing the KL divergence here means optimizing the probability distribution parameters (μ and σ) to closely resemble that of the target distribution.

Variational Autoencoder Intuition

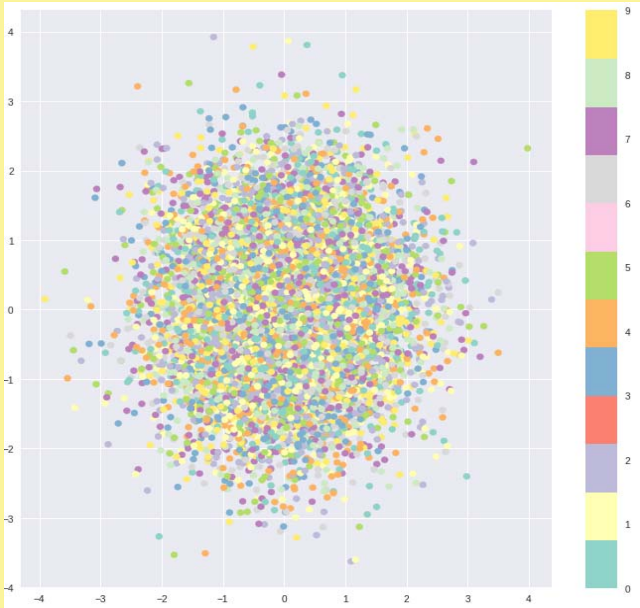
- In VAE, it is usually assumed that the distribution of the latent space follows a zero mean Normal distribution with diagonal covariance matrix (each component is independent of the other)
- KL divergence loss will encourage encodings from different inputs to be clustered about the center of the latent space
- If network creates clusters in specific regions then KL divergence loss will penalize such clusters formation

Variational Autoencoder Intuition

- But, only KL loss results in a latent space encodings densely placed randomly, near the center of the target distribution, with little regard for similarity/dis-similarity of input samples.

- The decoder finds it impossible to decode anything meaningful from this space, simply because there really isn't any structure/context specific meaning.

Variational Autoencoder Intuition



Latent space after training on MNIST when only optimized with KL loss

Image Source: https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

Variational Autoencoder Intuition

- Optimizing reconstruction loss + KL divergence loss results in the generation of a latent space which maintains the similarity of nearby encodings on the local scale via clustering
- Yet globally, is very densely packed near the latent space origin

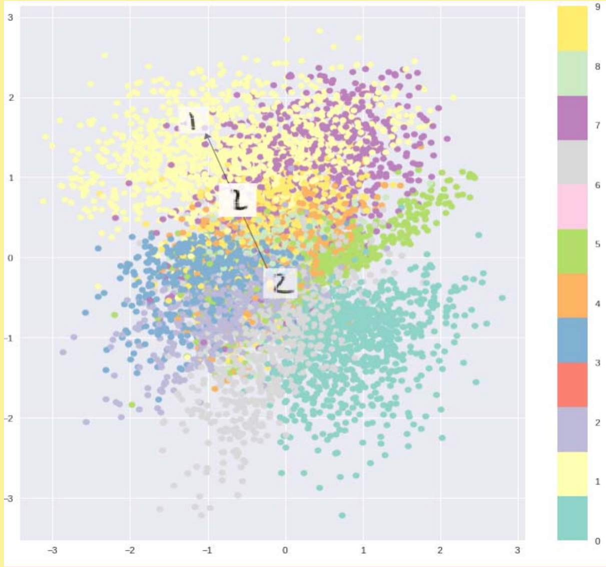
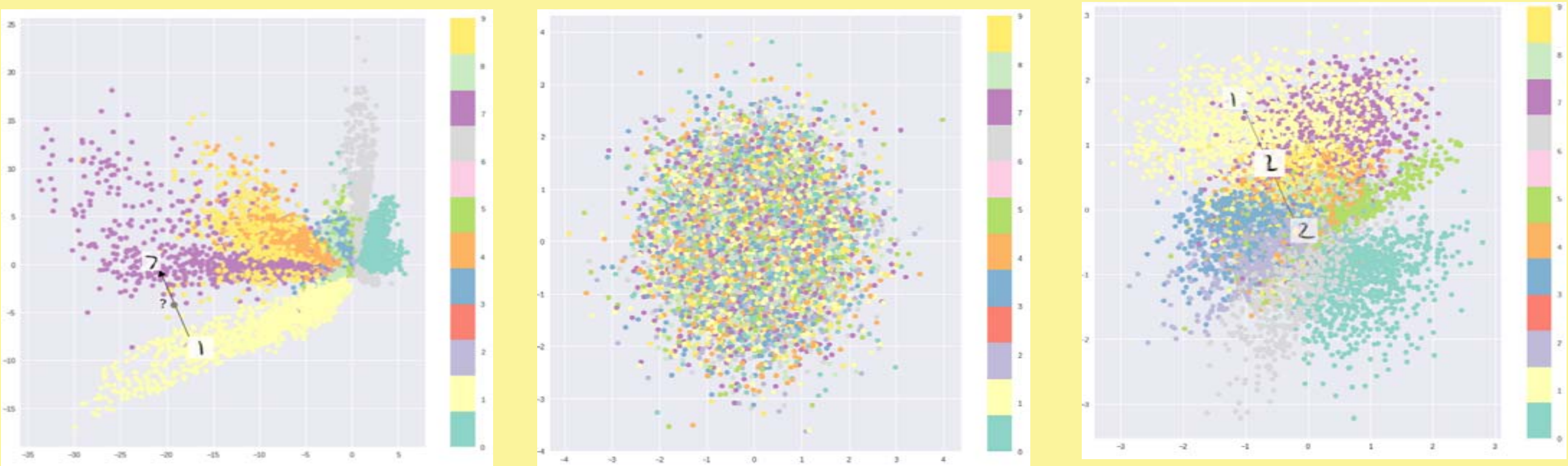


Image Source: https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

Variational Autoencoder Intuition



Reconstruction Loss | KL Divergence Loss | KL Divergence + Reconstruction Loss

Image Source: https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf


NPTEL ONLINE CERTIFICATION COURSES
Thank you