Normalizing Flows (Part 1)

Dr. Alireza Aghamohammadi

Modeling 1D with Normalizing Flows

Normalizing flows transform a simple distribution into a more complex one using a deep network to learn a probability model.

To model a 1D distribution $Pr(x)$:

- ❖ Start with a simple base distribution $Pr(z)$ over a latent variable z.
- ❖ Apply a function $x = f[z, \phi]$, where ϕ are parameters chosen to ensure $Pr(x)$ matches the desired distribution.

Generating a new example $x^\star$:

- ❖ Draw $z^\star$ from the base distribution.
- ❖ Pass $z^\star$ through the function: $x^\star = f[z^\star, \phi]$.

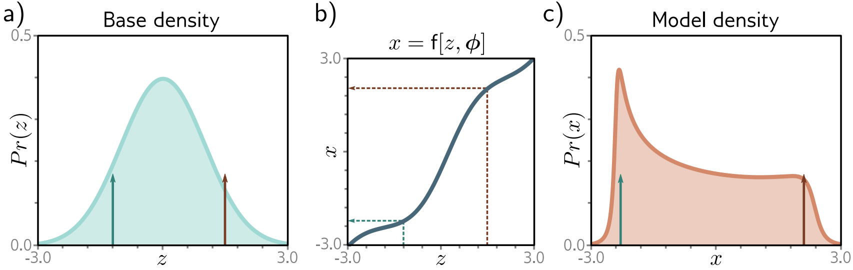

Image credit: Understanding Deep Learning

Measuring Probability of Data Point x

Measuring the probability of a data point x can be challenging.
Consider applying a function $f[z, \phi]$ to a random variable z with known density $P_r(z)$.
The probability of data x under the transformed distribution is given by:

$$P_r(x \mid \phi) = \left| \frac{\partial f[z, \phi]}{\partial z} \right|^{-1} \cdot P_r(z)$$

where $z = f^{-1}[x, \phi]$ is the latent variable that generated x.

Proof: To conserve density volume so that the distribution sums to one:

$$P_r(x \mid \phi)\,|dx| = P_r(z)\,|dz|$$

$$P_r(x \mid \phi) = P_r(z) \left| \frac{dz}{dx} \right|$$

$$P_r(x \mid \phi) = P_r(z) \left| \frac{\partial z}{\partial f[z, \phi]} \right|$$

$$P_r(x \mid \phi) = P_r(z) \cdot \left| \frac{\partial f[z, \phi]}{\partial z} \right|^{-1}$$

Normalizing Flows

Normalizing flows transform data in both directions.

- ❖ Forward Mapping: Also known as the generation direction.
- ❖ Base Density: Typically chosen as a standard normal distribution.
- ❖ Inverse Mapping: Known as the normalizing direction, it transforms the complex distribution over x into a normal distribution.

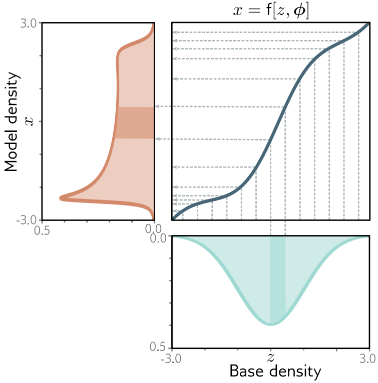

Image credit: Understanding Deep Learning

Learning Parameters

To learn the distribution, we find parameters ϕ that maximize the likelihood of the training data.

$$
\begin{aligned}
\hat{\phi} &= \underset{\phi}{\mathsf{argmax}} \left[\prod_{i=1}^{I} P_r(x_i \mid \phi) \right] \\
&= \underset{\phi}{\mathsf{argmin}} \left[-\sum_{i=1}^{I} \log \left[P_r(x_i \mid \phi) \right] \right] \\
&= \underset{\phi}{\mathsf{argmin}} \left[\sum_{i=1}^{I} \log \left| \frac{\partial f[z_i, \phi]}{\partial z_i} \right| - \log \left[P_r(z_i) \right] \right]
\end{aligned}
$$