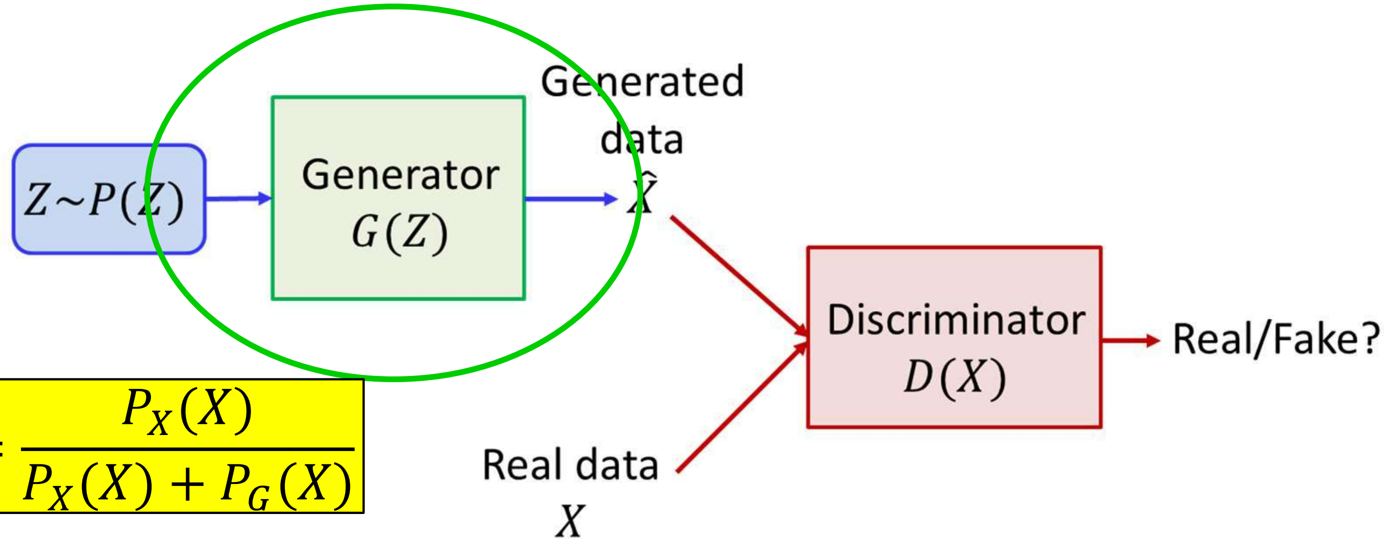
Analysis of optimal behavior:
The optimal generator



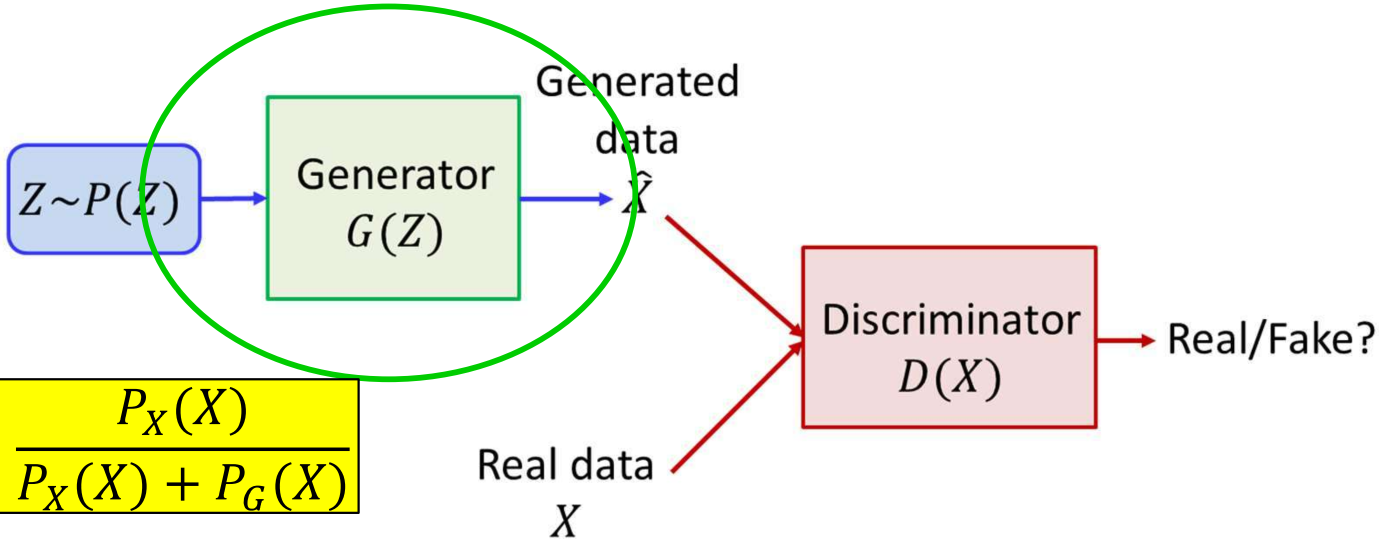
$$D(X) = \frac{P_X(X)}{P_X(X) + P_G(X)}$$

$$\min_G \max_D E_X \log D(X) + E_Z \log(1 - D(G(Z)))$$

- **With a perfect discriminator**:

$$L = E_{X \sim P_X(X)} \log D(X) + E_{X \sim P_G(X)} \log(1 - D(X))$$

Analysis of optimal behavior:
The optimal generator



$$D(X) = \frac{P_X(X)}{P_X(X) + P_G(X)}$$

$$\min_G \max_D E_X \log D(X) + E_Z \log(1 - D(G(Z)))$$

- **With a perfect discriminator**:

$$L = E_{X \sim P_X(X)} \log D(X) + E_{X \sim P_G(X)} \log(1 - D(X))$$

$$= E_{X \sim P_X(X)} \log\left(\frac{P_X(X)}{P_X(X) + P_G(X)}\right) + E_{X \sim P_G(X)} \log\left(\frac{P_G(X)}{P_X(X) + P_G(X)}\right)$$

The KL Divergence

$$KL\,(P,Q) = \sum_X P(X)\log(P(X)/Q(X))$$

- What are the problems with this?

$$KL\,(Q,P) = \sum_X Q(X)\log(Q(X)/P(X))$$

- What are the problems with this?

The KL Divergence

$$KL\,(P,Q) = \sum_X P(X)\log(P(X)/Q(X))$$

- What are the problems with this?

$$KL\,(Q,P) = \sum_X Q(X)\log(Q(X)/P(X))$$

- What are the problems with this?

KL is not symmetric, and runs into issues if either P or Q become 0 (whichever is inside the log)

The Jensen Shannon Divergence

$$JSD\ (P, Q)$$
$$= \ 0.5\ KL(P, 0.5(P+Q)) \ + \ 0.5KL(Q, 0.5(P+Q))$$

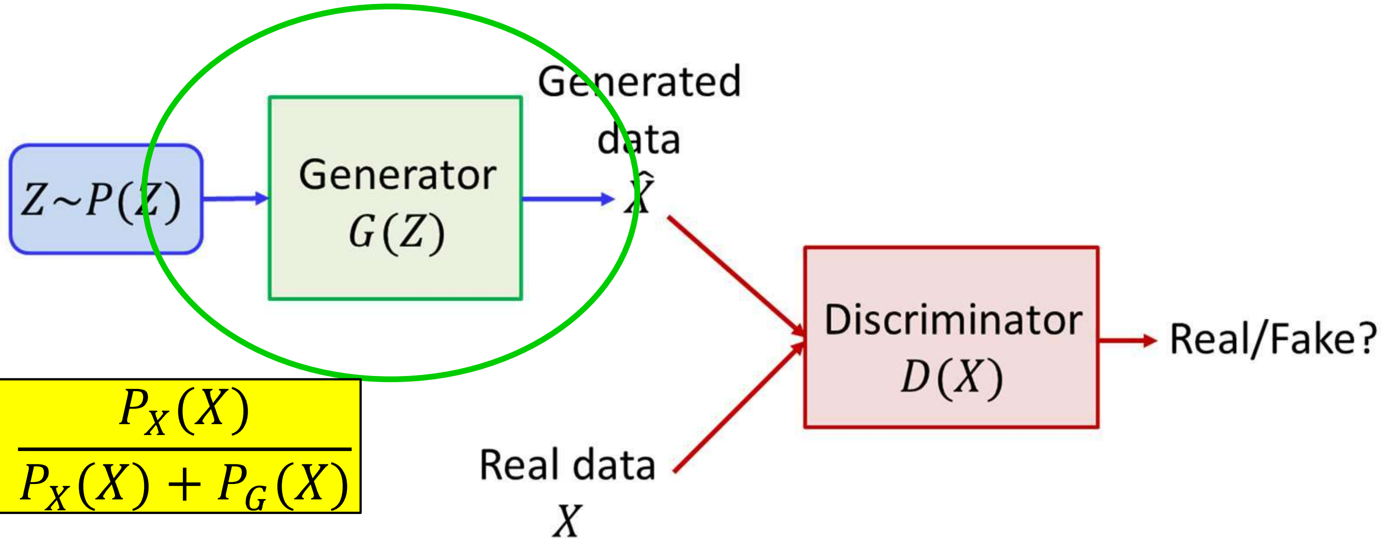
- If the term inside the log is 0, both P and Q are 0
  - 0 log 0 = 0, so there are no problems
- Also, this is symmetric: JSD(P,Q) = JSD(Q,P)

The Jensen Shannon Divergence

$$JSD\ (P, Q)$$
$$= 0.5\ KL(P, 0.5(P + Q)) + 0.5KL(Q, 0.5(P + Q))$$

- A symmetric variant of KL that does not exaggerate instances to which one of the distributions assigns 0 probability
  - $KL(P, Q) = \sum_X P(X) \log(P(X)/Q(X))$ blows up the contributions of X with $Q(X) = 0$

Analysis of optimal behavior: The optimal generator



$$D(X) = \frac{P_X(X)}{P_X(X) + P_G(X)}$$

$$\min_G \max_D E_X \log D(X) + E_Z \log(1 - D(G(Z)))$$

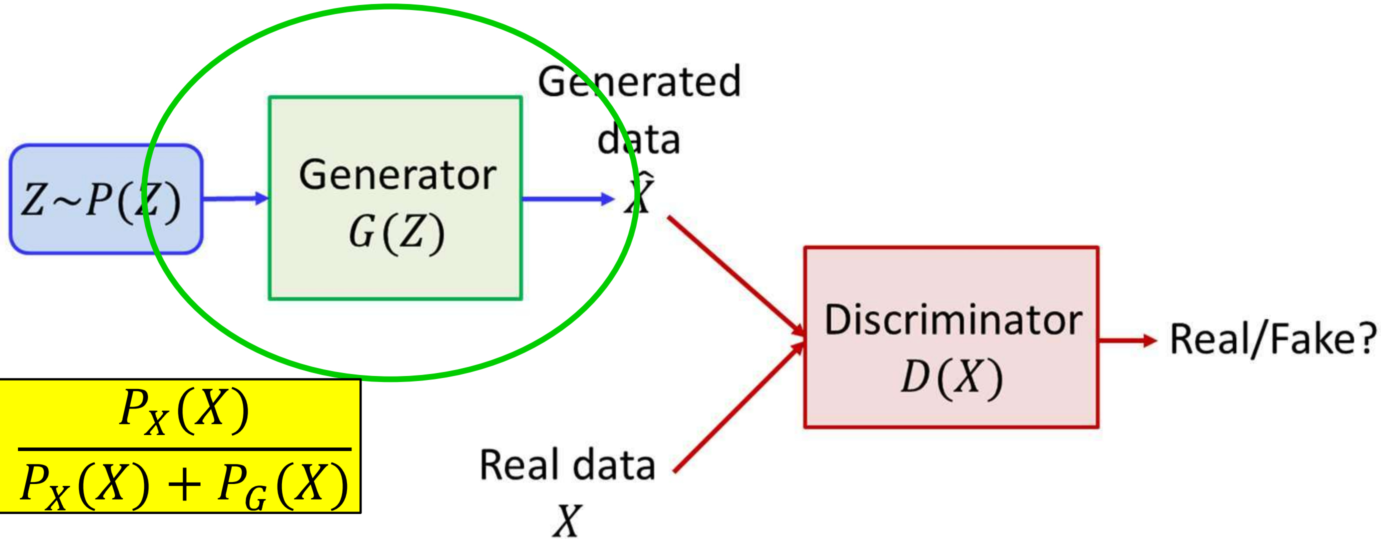
- **With a perfect discriminator**:

$$L = E_{X \sim P_X(X)} \log D(X) + E_{X \sim P_G(X)} \log(1 - D(X))$$
$$= E_{X \sim P_X(X)} \log\left(\frac{P_X(X)}{P_X(X) + P_G(X)}\right) + E_{X \sim P_G(X)} \log\left(\frac{P_G(X)}{P_X(X) + P_G(X)}\right)$$

- This is just the Jensen-Shannon divergence between $P_X(X)$ and $P_G(X)$ to within a scaling factor and a constant

$$L = 2JSD\big(P_X(X), P_D(X)\big) - \log 4$$

Analysis of optimal behavior:
The optimal generator



$$D(X) = \frac{P_X(X)}{P_X(X) + P_G(X)}$$

- The optimal generator:

$$\min_G 2JSD\big(P_X(X), P_G(X)\big) - \log 4$$

- The optimal generator minimizes the Jensen Shannon divergence between the distributions of the actual and synthetic data!
  - Tries to make the two distributions maximally similar