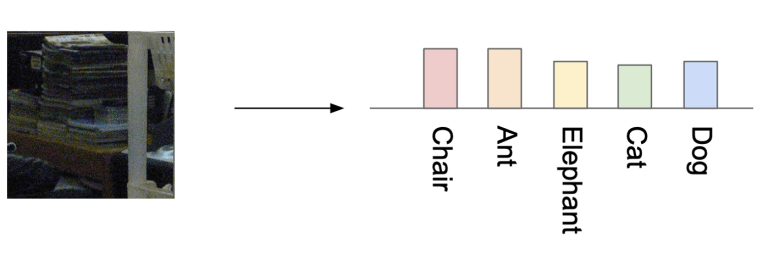
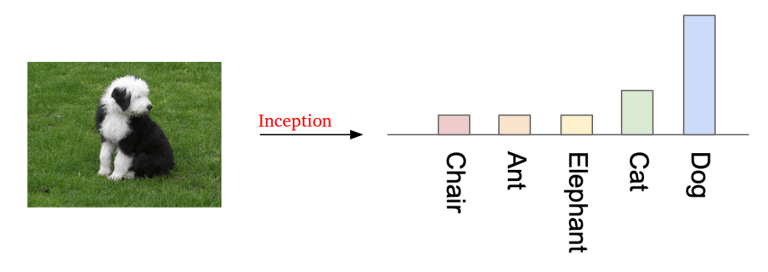
# GAN evaluation

Mọi người để ý là để đánh giá ảnh sinh ra mình cần dựa trên 2 yếu tố:

* Chất lượng ảnh: Mình muốn chất lượng ảnh cao, giống với dataset
* Độ đa dạng: Generator cần sinh ra được nhiều ảnh khác nhau khi mình random input z. Nếu Generator mãi chỉ sinh ra được một vài ảnh thì cũng không thực sự ý nghĩa lắm.

1. Inception Score (IS)

* Inception NN: input là 1 ảnh, output với hàm softmax sẽ cho ra được xác suất ảnh đấy thuộc từng lớp tương ứng (đoạt giải nhì cuộc thi *ImageNetChallange 2015*)



* Như các bạn có thể thấy hình bên trên:
* Input là ảnh 1 chú chó, đi qua mạng Inception sẽ cho ra xác suất là chó.
* Input là ảnh không thuộc lớp nào trong class, sẽ cho ra xác suất như bên phải(uniform).
* Dựa vào trên, ngta đưa ra được cách như sau:
* Mình có thể cho ảnh sinh ra qua mạng Inception nếu ảnh rõ, tốt thì mạng sẽ phân loại tốt (xác suất 1 lớp cao hơn hẳn) => kiểm tra được chất lượng ảnh.
* Cộng theo từng lớp các giá trị xác suất của tất cả các ảnh sinh ra trong generator lại. Nếu Generator có thể sinh ra đa dạng dữ liệu thì tổng xác suất sẽ dạng *uniform*, ngược lại nếu dữ liệu chỉ sinh ra dữ liệu ở 1 hay 2 lớp thì tổng xác suất sẽ chỉ cao hơn ở 1 hay 2 lớp.
* Inception Score (IS)
* Sử dùng Kullback–Leibler divergence (KLD) để đánh giá. Hiểu đơn giản thì Kullback–Leibler (KL) nhận input là 2 distribution và output ra độ tương đồng giữa hai distribution. Nếu 2 distribution giống nhau thì KL = 0, hai distribution càng khác nhau thì KL càng lớn.
* Như vậy, giá trị KL cao khi ảnh chất lượng cao và ảnh đa dạng. Ta có thể tính giá trị KL cho mỗi ảnh sinh ra rồi lấy trung bình lại làm giá trị IS cho model.



Trong đó p(y|x) là label distribution, p(y) là marginal distribution. Hàm exp để kết quả dễ so sánh hơn.

* Nhược điểm
* Vì mình dùng Inception pre-trained model với Imagenet dataset, nên nếu ảnh Generator sinh ra không thuộc lớp trong Imagenet thì giá trị output của model sẽ không ý nghĩa lắm => KL thấp.
* Nếu Generator chỉ sinh được một ảnh mỗi lớp thì chỉ số KL vẫn có thể cao => Chỉ đa dạng lớp nhưng không đa dạng ảnh trong mỗi lớp
* Nếu Generator nhớ và sinh ra các ảnh trong dataset thì chỉ số KL cũng cao => Không tốt.

1. Fréchet Inception Distance (FID)
2. Chỉ số ước lượng sự tương đồng cấu trúc (SSIM – Structural Similarity Index Measurement)
3. Tỷ số tín hiệu lớn nhất/ nhiễu (PSNR)