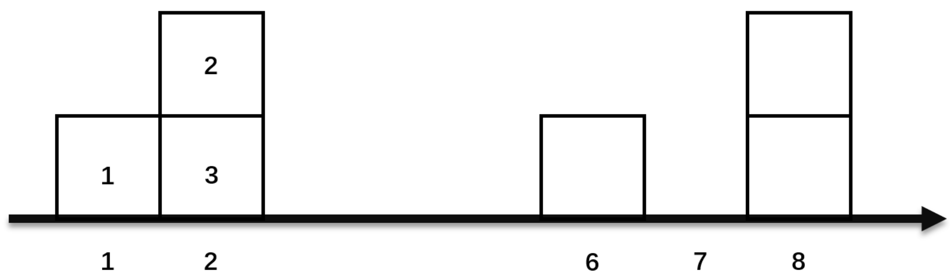
Quiz 10 **Name:** ______________________
Week 13, Dec/2/2020 **On your left:** ______________________
CS 280: Fall 2020 **On your right:** ______________________
Instructor: Xuming He, Lan Xu

Instructions:
Please answer the questions below. Show all your work. This is an open-book test. NO discussion/collaboration is allowed.

Problem 1. (10 points) **EM distance and WGAN**. According to the given figure, plot all possible earth transportation plan and compute the Earth-mover distance for each plan. What are the problems of Vanilla GAN and how could Wasserstein GAN solve them, please give detailed explanation. (You can find information from Lecture20 Page 12-26.)



Problem 2. (10 points) **Evaluation of GAN**

1. (1') What's the discriminator and generator loss curve like for a DCGAN if it is well trained. (2') If the loss for the discriminator has gone to zero or close to zero while the loss of the generator also rise and continue to rise over the same period, what's the possible reason? (2') If the DCGAN converges, does it means that it must be a well trained GAN? If not, what will the generator output in such a state?

2. List two widely adopted quantitative evaluation metrics for GAN, give their equations (and explain the meaning of the variables you used)(2x2'), what does a high or low score mean(1')?

Solution:

1. (a) The generator and discriminator are competing against each other, hence improvement on the one means the higher loss on the other, if a GAN converges, both discriminator and generator losses are converging to some permanent numbers.

   (b) Sometimes the loss for the discriminator has gone to zero or close to zero while the loss of the generator also rise and continue to rise over the same period. This type of loss is most commonly caused by the generator outputting garbage images that the discriminator can easily identify.

   (c) Even though the model may converges well, it still may suffer mode collapse problem, while a generator is only capable of generating one or a small subset of different outcomes, or modes.

2. Two widely adopted metrics for evaluating generated images are the Inception Score and the Frechet Inception Distance.

   (a) Inception Score.

   $$IS(G) = \exp(E_{\mathbf{x}}[D_{KL}(p(y|\mathbf{x})||p(y))])$$

   Given a generated sample $\mathbf{x}$, the Inception v3 network output a class label y, and $p(y|\mathbf{x})$ is the condition class distribution, $p(y)$ is the marginal distribution.

   The authors who proposed the IS aimed to codify two desirable qualities of a generative model into a metric:

   the Inception Network should be highly confident there is a single object in the image; The generative algorithm should output a high diversity of images from all the different classes in ImageNet;

   If both of these traits are satisfied by a generative model, then we expect a large KL-divergence between the distributions $p(y)$ and $p(y|\mathbf{x})$, resulting in a large IS.

   (b) Frechet Inception Distance

   $$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r\Sigma_g})$$

   FID is the Frechet distance between two multidimensional Gaussian distributions: $\mathcal{N}(\mu_r, \Sigma_r)$, the distribution of the Inception v3 features from the 'real' images used to train the GAN and $\mathcal{N}(\mu_g, \Sigma_g)$, the distribution of the Inception v3 features of the images generated by the GAN. μ and Σ are the sample mean and sample covariance matrices of the activations of the final layer of the pretrained Inception V3 network. The smaller the FID value, the closer the two distributions.