

Solutions to End-Semester Examination

Advances in Deep Learning (ECE/CSE 677)

- Q1:** i, iii. Self attention without positional encoding is permutation equivariant, can model pairwise token interactions.
- Q2:** D. Full self attention in a ViT layer has computational complexity of $O(N^2d)$.
- Q3:** A, C. ViT uses patch embeddings followed by Transformer layers and usually requires large scale pretraining to match CNN performance.
- Q4:** A, C, D. PGD adversarial training generates adversarial samples in each iteration, logit alignment can improve robustness, and adversarial samples may lie off the clean data manifold.
- Q5:** A. Temperature smooths the teacher logits which reveals additional information about non dominant classes.
- Q6:** B, D. The student is trained using both soft teacher targets and true labels, and distillation allows compression of large models into smaller ones.
- Q7:** B. MAML learns an initialization that can be adapted quickly to new tasks using only a few gradient steps.
- Q8:** A, B, C, D. All listed steps are part of the MAML meta training procedure for one shot N way tasks.
- Q9:** A, B, D. The score is defined as $\nabla_x \log p(x)$, Langevin dynamics uses the score to move toward high density regions, and annealed Langevin dynamics applies multiple noise levels during sampling.
- Q10:** B. A standard GCN layer uses normalized adjacency with self loops resulting in $\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} \Theta^{(l)}$.
- Q11:** A, B, C. The Inception Score increases when generated images are clearly classifiable and diverse across classes which corresponds to low conditional entropy and high marginal entropy.
- Q12:** A, B, D. FID compares Gaussian feature distributions, remains unchanged under invertible linear feature transformations, and reflects both quality and diversity of generated images.
- Q13:** A, B, C, D. All statements correctly describe aspects of spectral graph convolution and its approximations including Chebyshev polynomials and the GCN first order simplification.
- Q14:** False. A standard GCN uses degree normalization and self loops rather than operating only on the raw adjacency matrix.