Module 8

The Future: Multimodal Models, Voting Ensembles, and Federated Learning

▼ What is transfer learning?

- The process of storing knowledge and experience gathered by one network and conferring it upon another network.
- This allows model builders to take advantage of commonalities and differences between different tasks, increase effectiveness of generalization, and reduce training requirements.

▼ What is multi-task learning?

- A form of transfer learning that intentionally teaches a model multiple tasks in a similar domain.
- This is possible through regularization.
- When you perform regularization when learning a task that is novel but related to previous tasks, it is superior at generalizing and avoiding overfitting.

▼ What is pre-training?

- When the model is trained on a much larger and more generalized dataset first, before being trained on more specific tasks in the same domain.
- This improves the accuracy of subsequent training on more specific training sets.

▼ If a dataset is small, what layers should be tuned when fine-tuning?

train only the final network layers to avoid overfitting to the new data

▼ What is metalearning?

- Metalearning, or "learning to learn", is a new technique that involves using learning algorithms on the metadata of a dataset in hopes of discovering the optimal way to converge on a solution for that data.
- Metalearning is used to give the developer insights that may assist in choosing methods of augmenting the dataset, selecting features, and tuning hyperparameters.

▼ What is federated learning?

- Federated learning, sometimes called collaborative learning, is a decentralized training technique wherein an algorithm is trained in multiple pieces, usually edge devices.
- This technique makes use of transfer learning methods to gather experience that is more variable and localized. Data is not shared between models during training, making it decentralized.
- Federated learning is different from TL in that each model is not identical. Rather, each model is unique and data is distributed unevenly before updating a copy of the model stored in a central server without sharing actual training data. This eliminates the need for a centralized data pool, reducing storage requirements and communication bandwidth.
- These models are then aggregated and prepared for inference.
- Federated learning also has implications for privacy, since training data does not need to be shared centrally.

▼ What are ensemble methods?

- Ensemble methods are another form of multi-model learning.
- In ensemble learning, multiple different models, which can be of entirely different types and with different hyperparameters, feature sets, and learning methods, are used together toward a common task or series of tasks in a common domain.
- Ensemble methods can also be used cross-domain.

- The important factor in building a working ensemble learning model is a **commonality of outputs**. This can mean that one model is used to tune hyperparameters for another model or pre-process data in some way for another model (you can use meta-learning for this). It could also mean using multiple models to process the same data in a different way, or process different data towards achieving the same goal.

▼ What is evolutionary (genetic algorithms) optimization for hyperparameters?

- When used for optimizing hyperparameters, evolutionary optimization utilizes evolutionary algorithms to look for an efficient algorithm by creating an initial population of possible hyperparameter variations, running and testing them, assigning each hyperparameter tuple a rank, and use crossover and mutation techniques to replace poorly performing tuples with newly generated ones.
- This process is repeated until accuracy and efficiency of the tested models is acceptable.

▼ What are cloud and edge technologies?

- Cloud technologies can be leveraged to enhance AI deliverability and performance.
- When combined with inference and even federated training on the edge can reap large functional gains.
- Cloud cluster processing frameworks such as Hadoop and Spark will enjoy increased future use, while more devices model on the edge at the same time.

▼ What are TPUs, ASICs, memristors, and neuromorphic chips?

- New specialized chipsets designed with machine learning in mind are already granting enormous performance improvements in both training and inference.
- ASICS can provide massive parallelism and performance boosts by using low-precision arithmetic.
- Google's Tensor Processing Units are specialized ASIC chipsets designed for tensor processing and vectorization.

- Neuromorphic chips are analog chips designed to mimic biological neural functions.
- Memristors are a form of transistor that doubles as a form of non-volatile memory, and can thus retain it's instructions.
- Chipsets based on memristors instead of transistors will allow more accurate modeling of Hebbian synaptic plasticity and neural spiking.

▼ What is Neuralink?

- Neuralink (Neuralink, n.d.) is a neural prosthetic designed to allow for two-way communication between the human brain and a computer system. This is known as a brain-machine interface (BMI).
- According to Neuralink, "Modulating neural activity will be an important part of next-generation clinical brain-machine interfaces...for example to provide a sense of touch or proprioception to neuroprosthetic movement control" and, while being only tangential to AI and ML, provides a potential for much greater understanding of biological neural networks and their functions.
- It also opens up a new source of data for AI training in the form of high-granularity neural spiking data.