

Prompts

Generative Models – In supervised learning we are interested in models that predict y based on x . Generative models have the ability to generate new x samples that are distinct from training data but are drawn from the same population. Popular generative applications that you may have heard about are models that produce photo-realistic images or music based on real but distinct training data (i.e., photos and music examples). In these cases, the x produced by the generative model are images and music, respectively. In a chemical application, a generative model would typically return new chemical structures with certain target properties. For example, imagine a model that could suggest new pharmaceutical molecules given existing molecules with a certain activity. This would be a big deal, and that is why a lot of people are working on it.

For your paper, provide an introduction and summary of what generative models are (~1pg) followed by a review of at least two journal papers (~2 pgs) that use or develop generative models for a chemical or engineering application. The journal articles that you review should have been published on or after **2019**, but your supporting references can be from anytime.

Active Machine Learning—In supervised learning we are concerned with training a model to predict y based on a x , using fixed training data. Active machine learning is interested in how to optimally designate and expand the training data to improve the predictions of a model. For example, say that we have a model that predicts whether a chemical plant is operating within specifications (similar to the Dow case) and we suspect that the model has bad performance for certain operating conditions. What set of data should we collect for training in order to improve the performance of the model? Active learning strategies address these kinds of challenges.

For your paper, provide an introduction and summary of what active learning is (~1pg) followed by a review of at least two journal papers (~2 pgs) that use or develop active learning approaches for a chemical or engineering application. The journal articles that you review should have been published on or after **2019**, but your supporting references can be from anytime.

Ensemble Methods—In the context of supervised learning, we have discussed individual models that can be used to predict the y given x , with linear regression being the simplest example. However, there is a family of methods known as “ensemble methods” where you actually train many independent models and combine their individual predictions to form a final prediction on y . Individually the models may not be very accurate, but as an ensemble their mean or mode predictions may be very accurate. Random forests are the most common example of ensemble methods, feel free to focus your paper on these or any other ensemble model that interests you.

For your paper, provide an introduction and summary of what ensemble models are (~1pg) followed by a review of at least two journal papers (~2 pgs) that use or develop ensemble models for a chemical or engineering application. The journal articles that you review should have been published on or after **2019**, but your supporting references can be from anytime.

Transfer Learning—In the context of supervised learning, we have assumed that we have enough data in order to train an effective model. But in many scenarios data is extremely scarce, and our ability to collect more data is likewise limited. Transfer learning is a subfield of machine learning that tries to utilize data from multiple sources and of different kinds in order to circumvent the limitations of data scarcity.

For your paper, provide an introduction and summary of what transfer learning is (~1pg) followed by a review of at least two journal papers (~2 pgs) that use or develop transfer learning methods for a chemical

or engineering application. The journal articles that you review should have been published on or after **2019**, but your supporting references can be from anytime.

Formatting Details: The paper should be single-spaced with one inch margins and employ 12pt Times New Roman font. The text should not exceed three pages and should have no more than four figures. A bibliography, formatted at your discretion, should be included at the end and does not count towards the page limit.

Due Date: May 3. Papers should be uploaded to Brightspace.

Additional notes:

- You are encouraged to copy figures from the papers that you review into your report.
- You can upload your report either as a pdf or word document.