medium.com

Cheat Sheets for Machine Learning Interview Topics - Aqeel Anwar - Medium

Aqeel Anwar

6–7 minutes

Illustration credit rawpixels.com

> **Updates:**
>
> Dec 25, 2021: Added Auto Encoder and variational Encoder
>
> Dec 25, 2020: Added Ensemble Methods

> Download the updated version of the cheat sheets from http://cheatsheets.aqeel-anwar.com/

A couple of years ago I started applying for internships in the area of Machine Learning and ML system design. I had been studying and actively researching in the area of ML for a few

years then. I was familiar with most of the basic topics. But when I started interviewing, I realized that though I had a general understanding of the topics, I required a quick go-through before I can answer it perfectly.

So I decided to refresh my concepts. I realized that before every interview, I was required to go through the topics again. So, I created my handwritten notes. Skimming through them was much easier than going through slides and book chapters. It provided me with a quick boost to my understanding in a short amount of time. I decided to convert my hand-written notes into compact cheat sheets that might come in handy for ML interviews and daily data-scientist life in general.

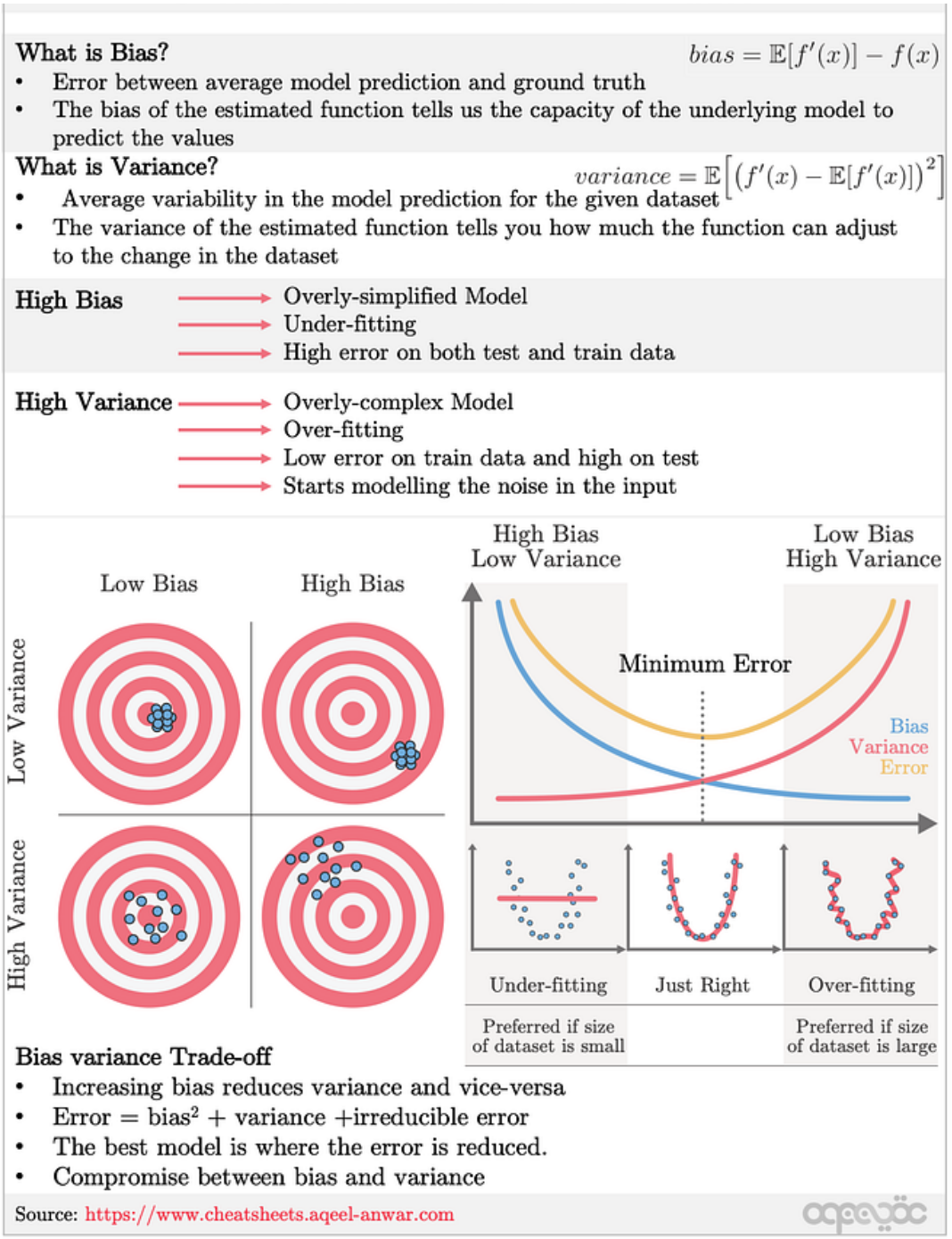
The rest of the article is based on those cheat sheets. For each topic, I provide

- An overview in form of a cheat sheet
- Example interview questions
- Suggested articles for a detailed understanding of the topic.

> **Note 1:** These cheat sheets are aimed at refreshing the concepts and are not meant to provide in-depth understandings of the topics for beginners.
>
> **Note 2:** The article is constantly updated for more cheat sheets.
>
> **Source:** All of these cheat sheets (and more) can be downloaded in pdf format from **www.cheatsheets.aqeel-anwar.com**.

Bias and Variance in Machine Learning Models

a) Overview:



Bias-variance tradeoff cheat sheet — Image by author

b) Example Questions:

1. What is Bias in ML models?
2. What is Variance in ML models?
3. What is the trade-off between bias and variance?
4. What are the demerits of a high bias / high variance ML model?
5. How do you select the model (high bias or high variance) based

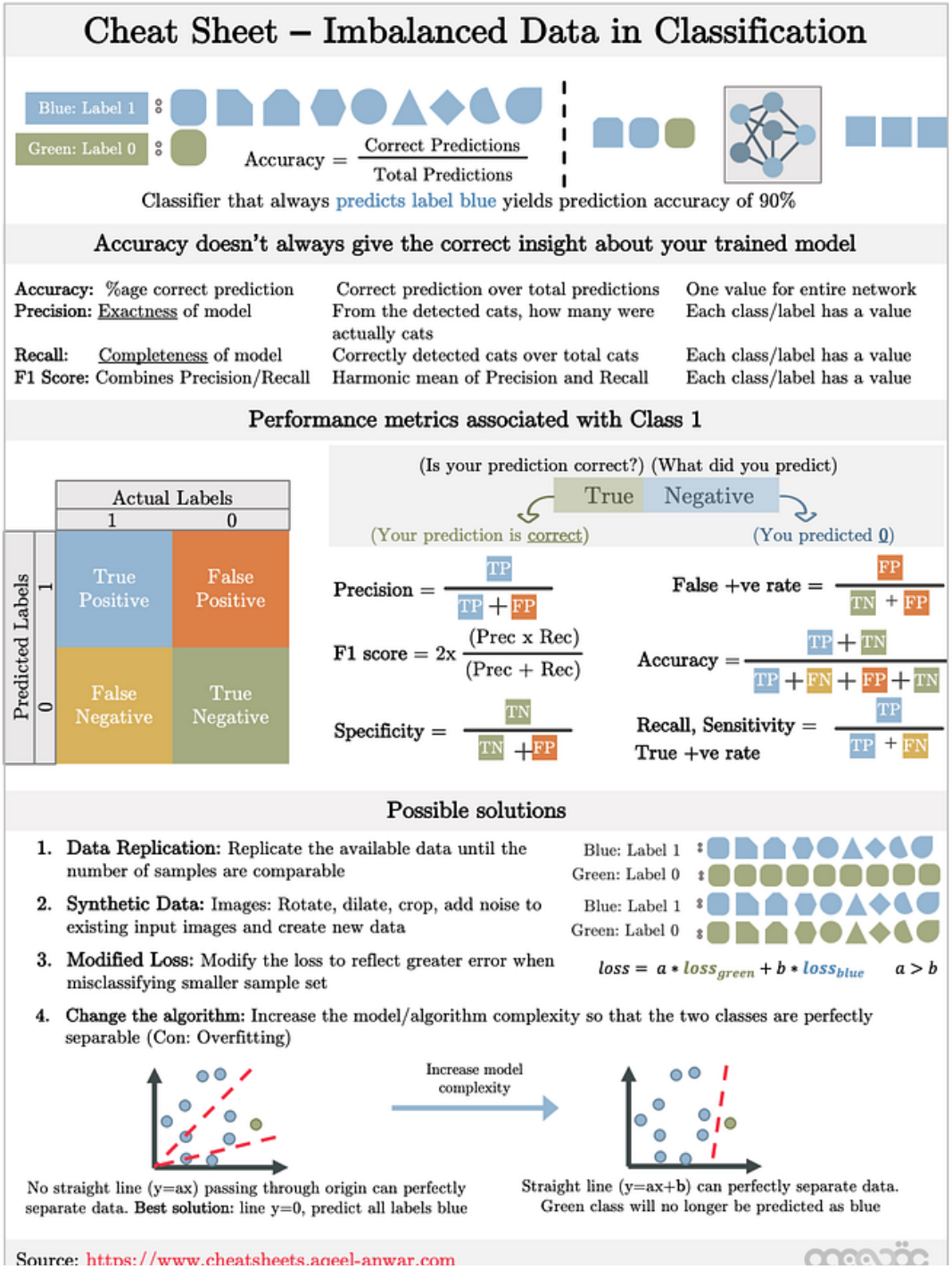
on the training data size?

c) Detailed Article:

- [Understanding the Bias-Variance Tradeoff and visualizing it with example and python code]

Imbalanced data in Machine Learning

a) Overview:



Imbalanced data in classification — Image by Author

b) Example Questions:

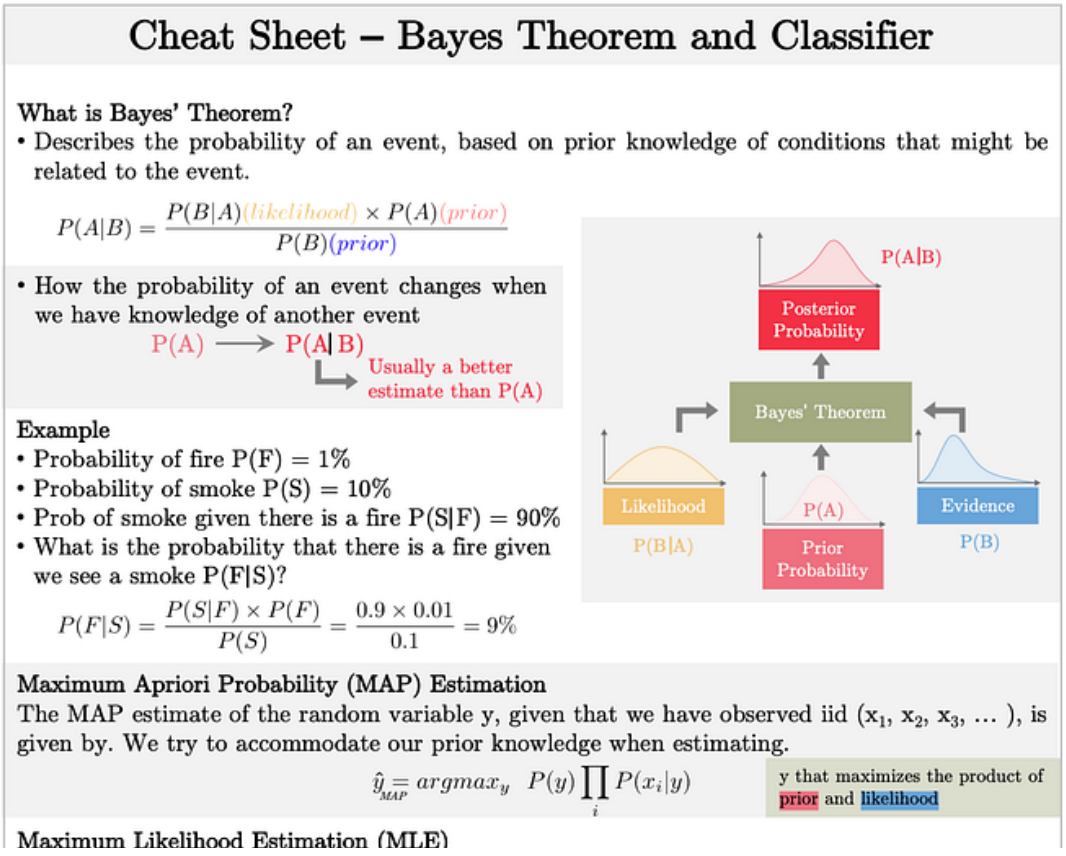
1. What is imbalanced data in classification?
2. Is accuracy a good performance metric? When does it fail to capture the performance of an ML system?
3. What are Precision and Recall? Give an example
4. How to address the issue of imbalanced data?

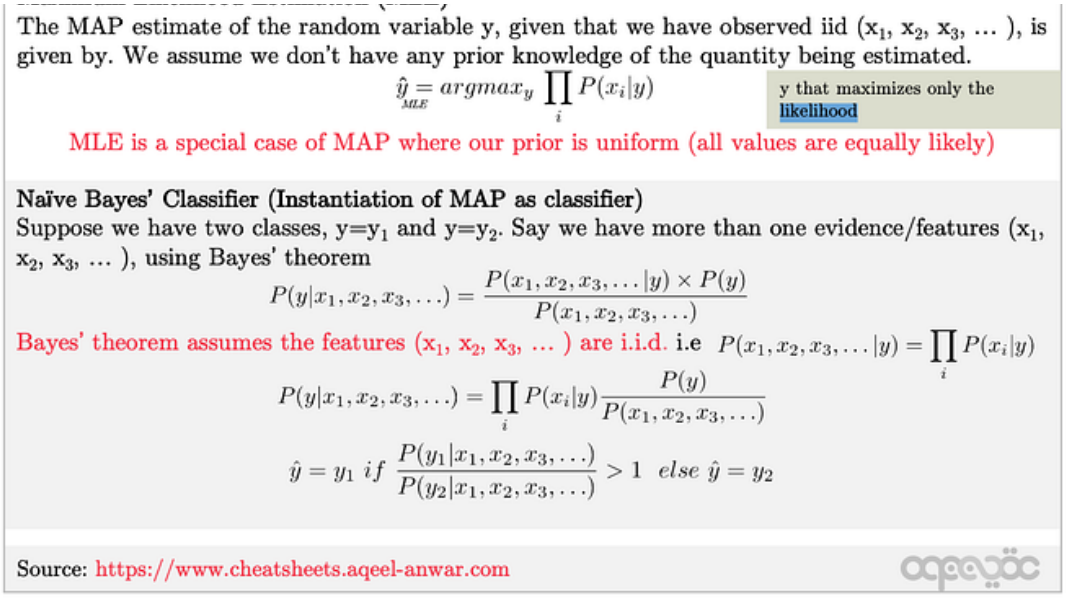
c) Detailed Articles:

- A walk through imbalanced classes in machine learning through a visual cheat sheet

Bayes' Theorem

a) Overview:





Bayes' Theorem and Classifier — Image by Author

b) Example Questions:

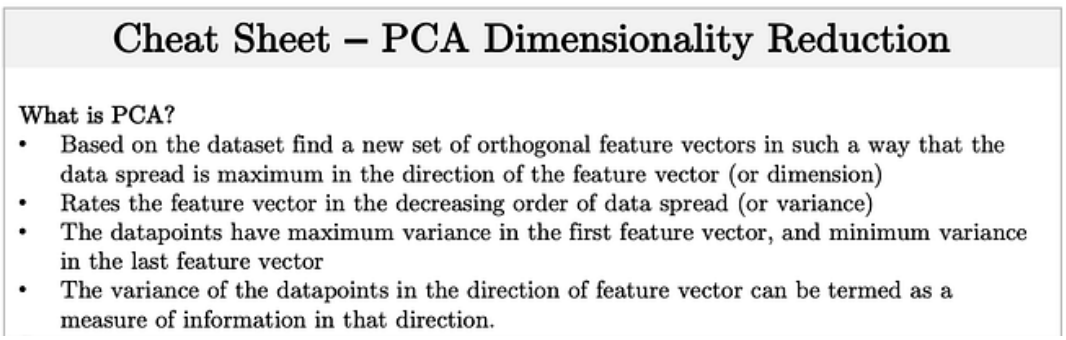
1. What is Bayes' theorem?
2. Toy example to implement Bayes' theorem
3. What is the difference between MLE and MAP?
4. When are MAP and MLE equal?

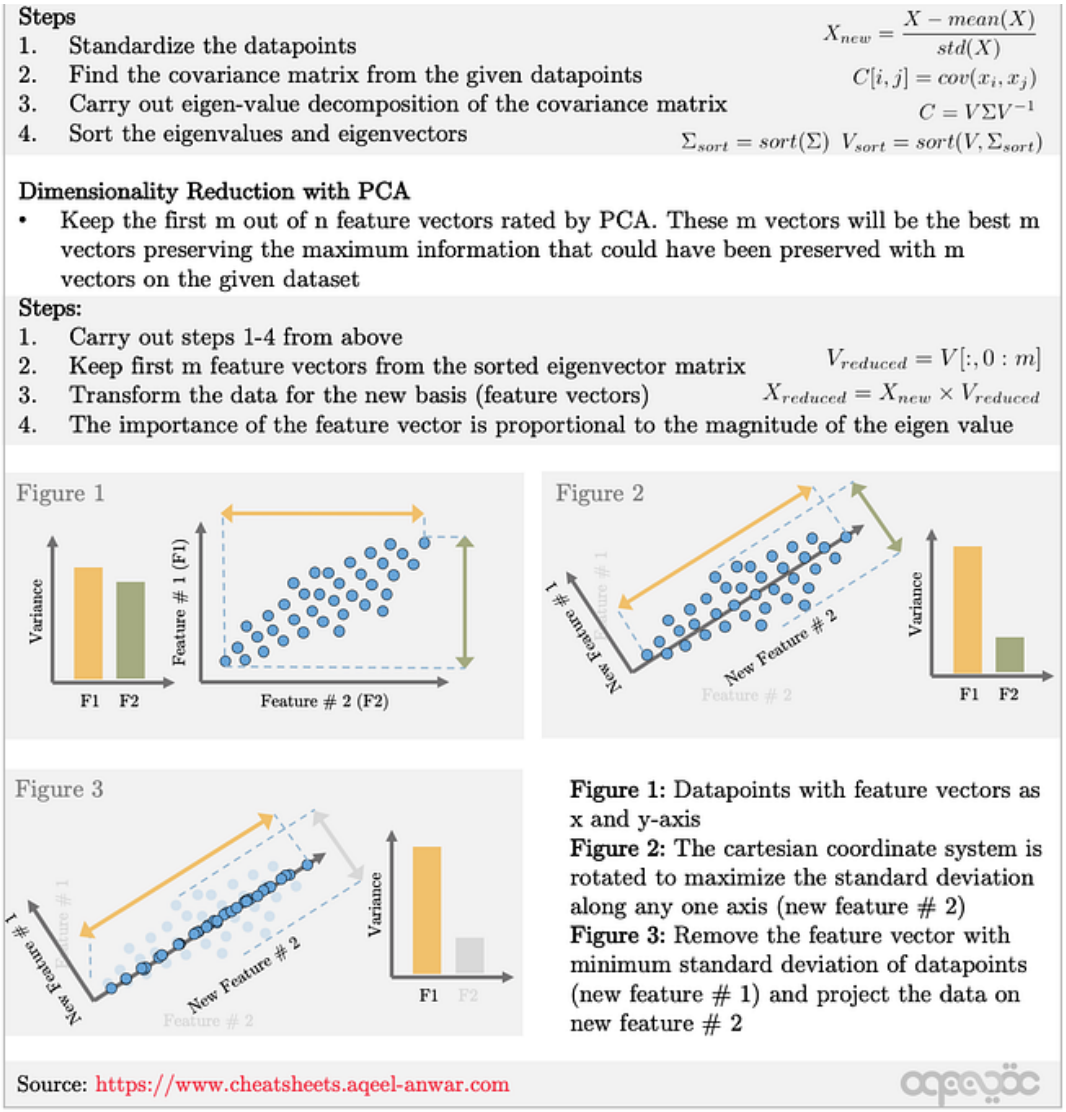
c) Detailed Articles:

- A Gentle Introduction to Bayes Theorem for Machine Learning

Principal Component Analysis and Dimensionality Reduction

a) Overview:





PCA and Dimensionality Reduction — Image by Author

b) Example Questions:

1. What is Principal Component Analysis?
2. How can we use PCA to reduce dimensions?
3. What do the eigenvalues signify in the context of PCA? *(Greater the magnitude of eigenvalue, the more information is preserved if we keep that corresponding eigenvector as a feature vector for our data)*

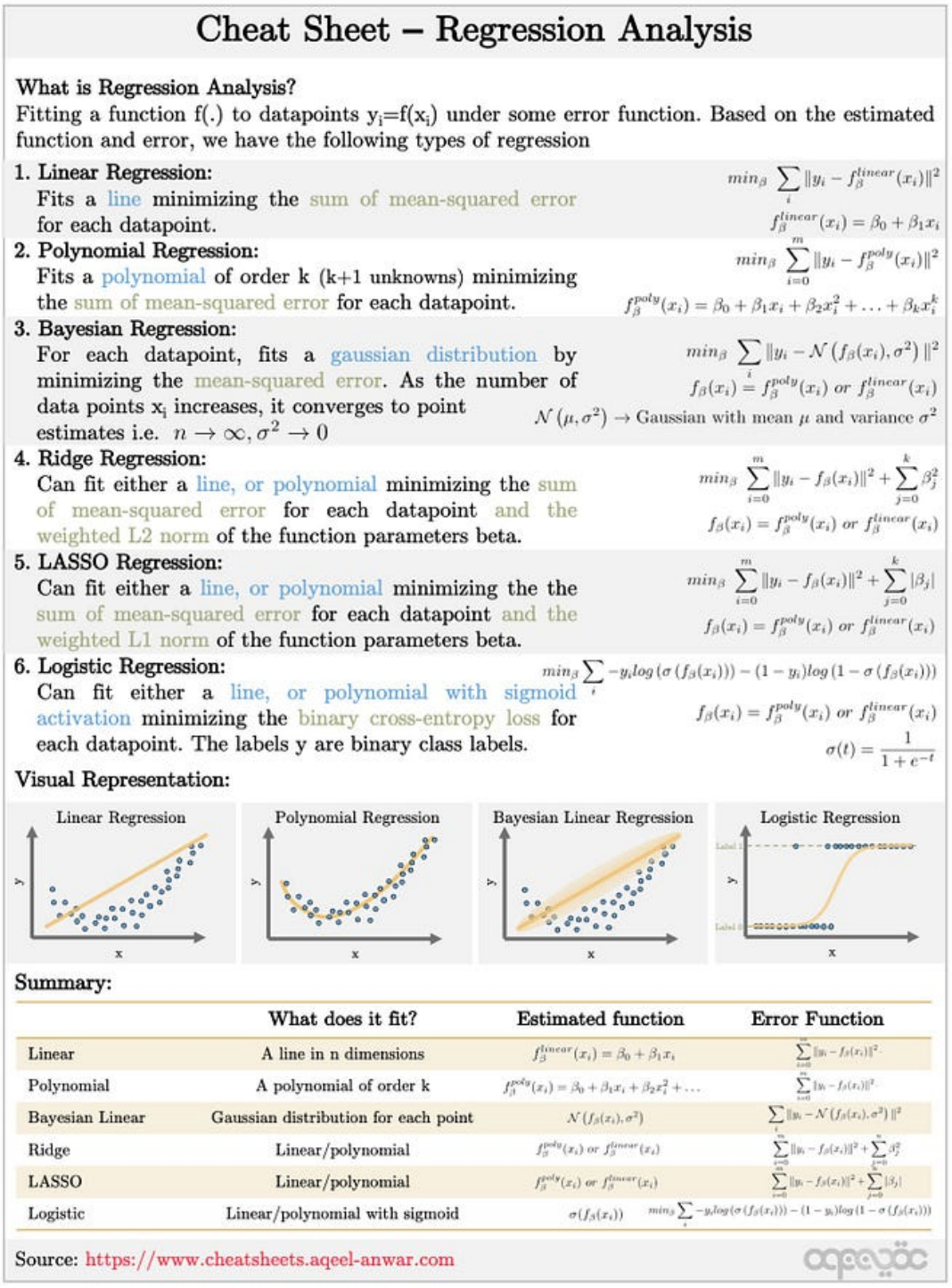
c) Detailed Articles:

- Understanding Principle Component Analysis(PCA) step by

step.

Regression in Machine Learning

a) Overview:



Regression Analysis — Image by Author

b) Example Questions:

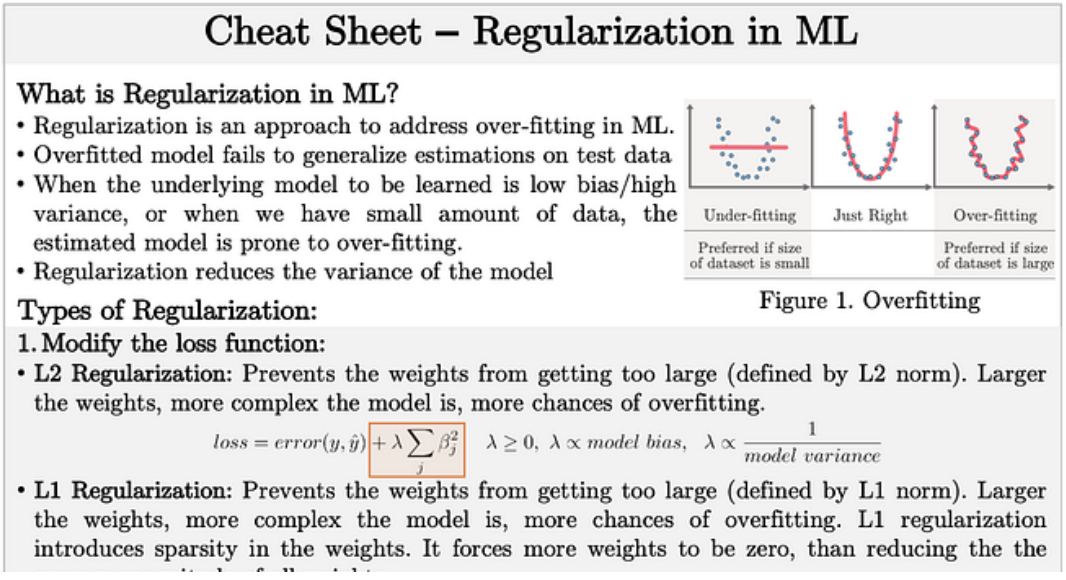
1. What is Regression in ML?
2. How can we introduce regularization in regression? *(LASSO and Ridge)*
3. What impact does LASSO and Ridge regression has on the weights of the model? *(Ridge tries to reduce the size of the weights learned, whereas LASSO tries to force them to zero creating a more sparse set of weights)*
4. When does the prediction by Bayesian linear regression approach the prediction of linear regression? *(When the number of data points is large enough)*
5. Is logistic regression a misnomer? *(Yes, because it is not regression, but classification based on regression)*

c) Detailed Articles:

- 5 Types of Regression and their properties
- Introduction to Bayesian Linear Regression

Regularization in Machine Learning

a) Overview:



Cheat Sheet – Regularization in ML

What is Regularization in ML?

- Regularization is an approach to address over-fitting in ML.
- Overfitted model fails to generalize estimations on test data
- When the underlying model to be learned is low bias/high variance, or when we have small amount of data, the estimated model is prone to over-fitting.
- Regularization reduces the variance of the model

Figure 1. Overfitting

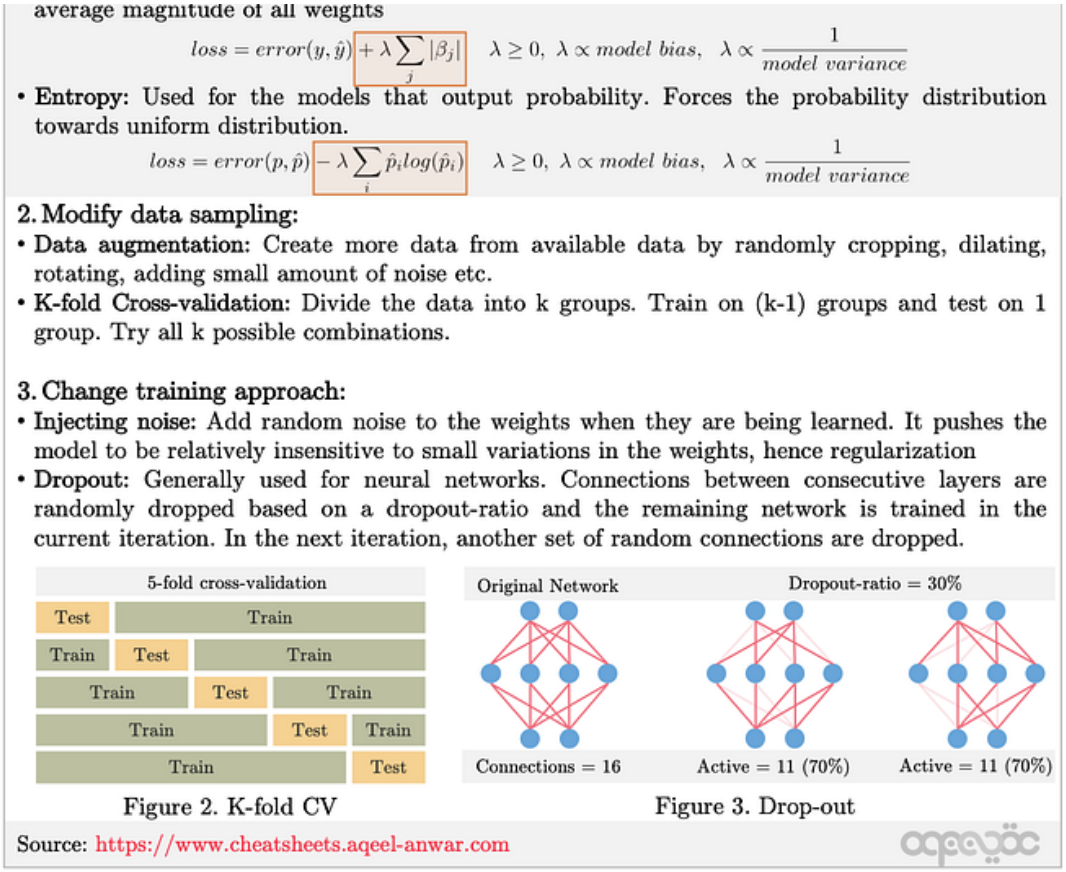
Types of Regularization:

1. Modify the loss function:

- **L2 Regularization:** Prevents the weights from getting too large (defined by L2 norm). Larger the weights, more complex the model is, more chances of overfitting.

$$loss = error(y, \hat{y}) + \lambda \sum_j \beta_j^2 \quad \lambda \geq 0,\ \lambda \propto model\ bias,\ \lambda \propto \frac{1}{model\ variance}$$

- **L1 Regularization:** Prevents the weights from getting too large (defined by L1 norm). Larger the weights, more complex the model is, more chances of overfitting. L1 regularization introduces sparsity in the weights. It forces more weights to be zero, than reducing the the



Regularization in ML — Image by Author

b) Example Questions:

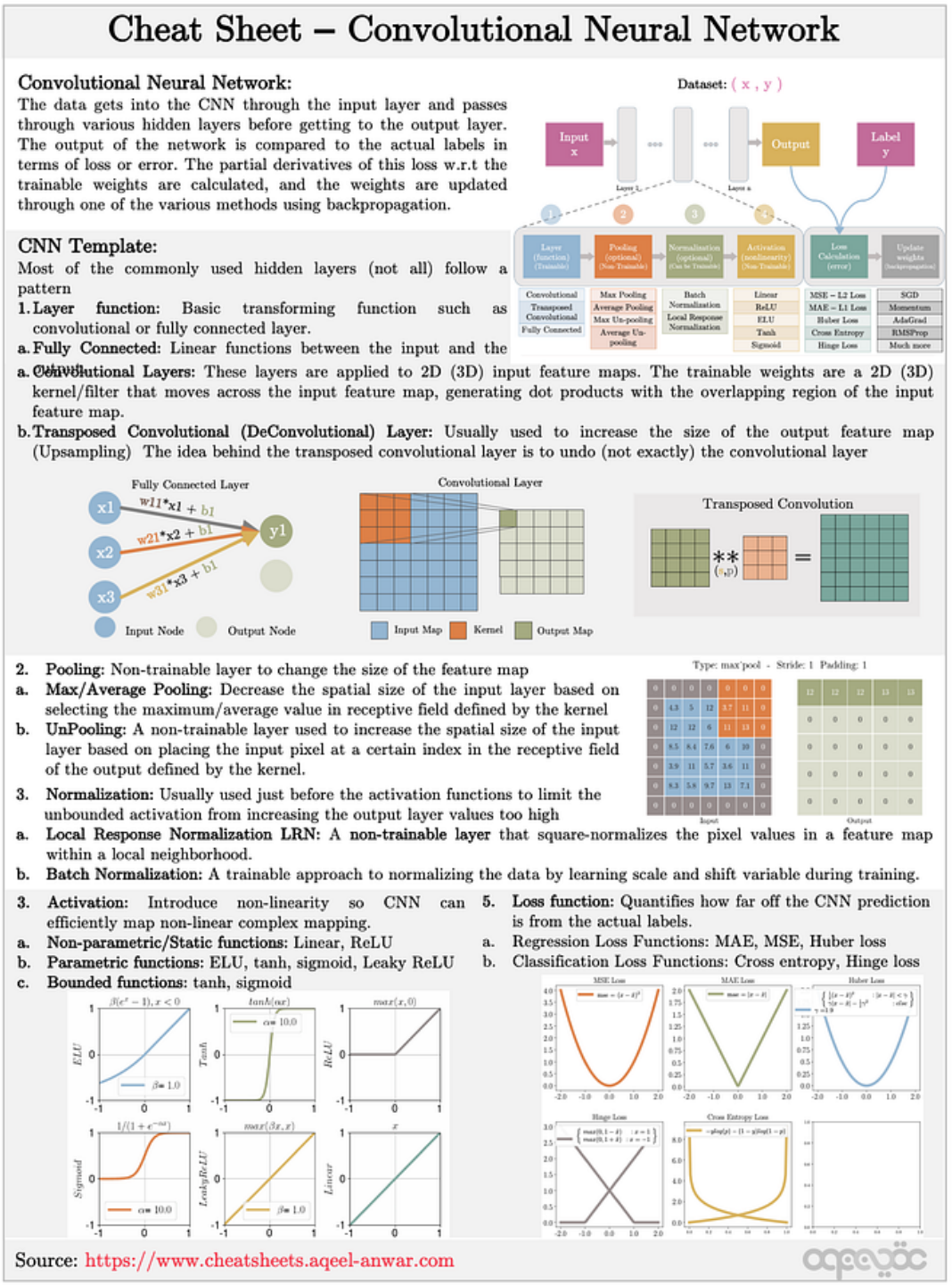
1. What is regularization in ML?
2. How can we address over-fitting?
3. What is K-fold cross-validation?
4. What is the difference between L1 and L2 regularization?
5. Why do we use dropout?

c) Detailed Articles:

- Regularization
- Regularization in Machine Learning

Basics of Convolutional Neural Network

a) Overview:



Convolutional Neural Network — Image by Author

b) Example Questions:

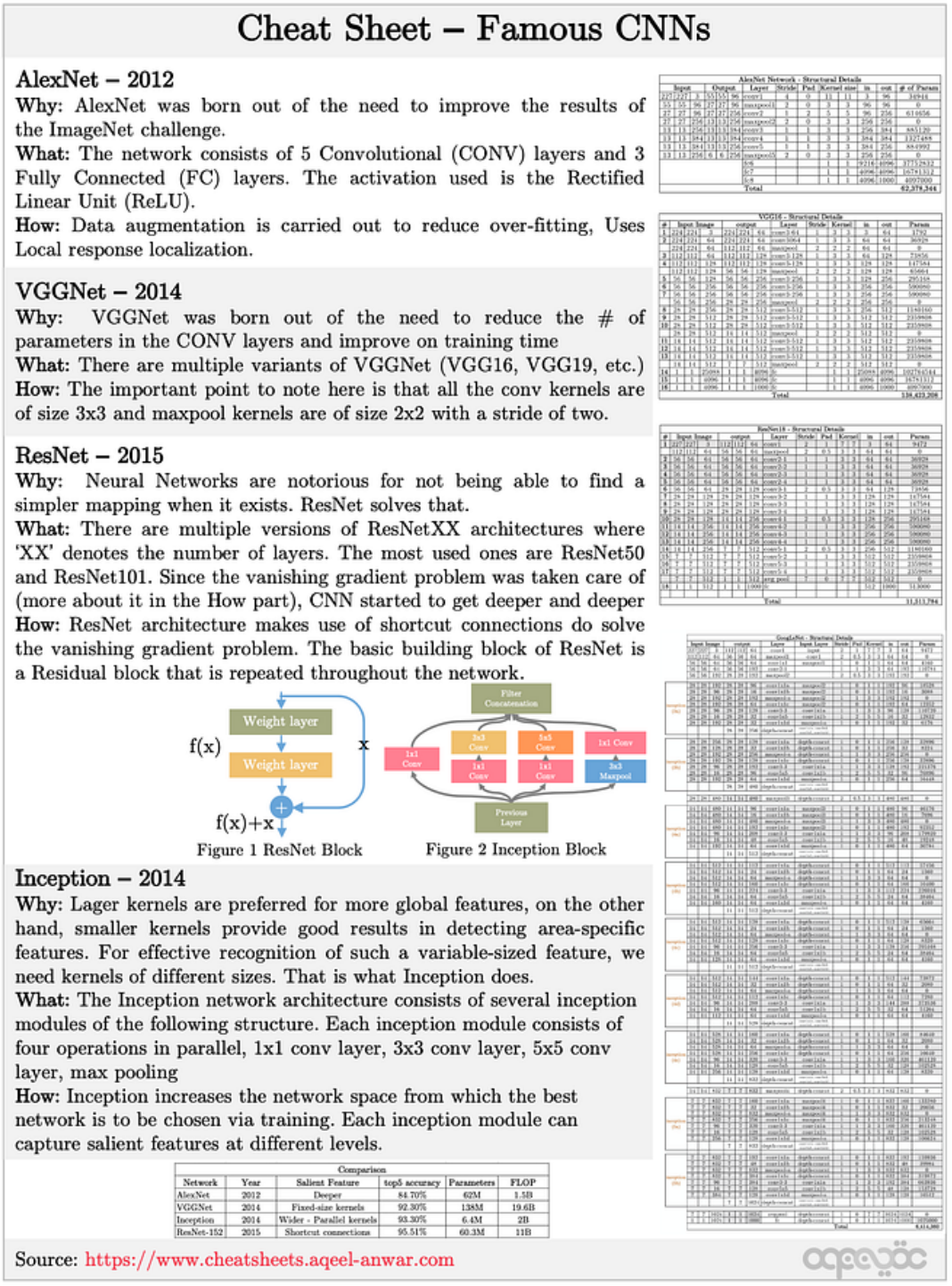
1. What is CNN?
2. Explain the difference between the convolutional layer and transposed convolutional layer.
3. What are some of the loss functions used for classification?

c) Detailed Article:

- What is a Convolutional Neural Network?

Famous DNNs in Machine Learning

a) Overview:



Famous CNNs — Image by Author

b) Example Questions:

1. How does the ResNet network address the problem of vanishing gradient?
2. What is one of the main key features of the Inception Network?
3. What are shortcut connections in the ResNet network?

c) Detailed Articles:

- Difference between AlexNet, VGGNet, ResNet, and Inception

Autoencoder and Variational Autoencoder

a) Overview:

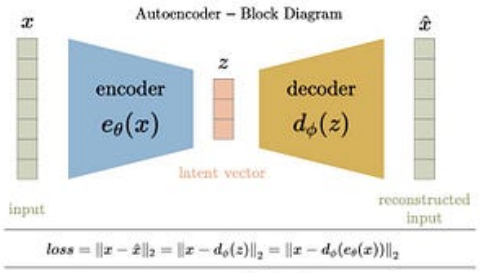
Cheat Sheet – Autoencoder & Variational Autoencoder

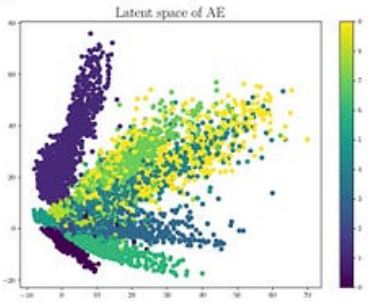
Context – Data Compression

- Data compression is an essential phase in training a network. The idea is to compress the data so that the same amount of information can be represented by fewer bits.

Auto Encoder (AE)

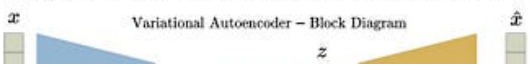
- Autoencoder is used to learn efficient embeddings of unlabeled data for a given network configuration. It consists of two parts, an encoder, and a decoder.
- The encoder compresses the data from a higher-dimensional space to a lower-dimensional space (also called the latent space),while the decoder converts the latent space back to higher-dimensional space.
- The entire encoder-decoder architecture is collectively trained on the loss function which encourages that the input is reconstructed at the output. Hence the loss function is the mean squared error between the encoder input and the decoder output.
- The latent variable is not regularized. Picking a random latent variable will generate garbage output.
- Latent variable is deterministic values and the space lacks the generative capability

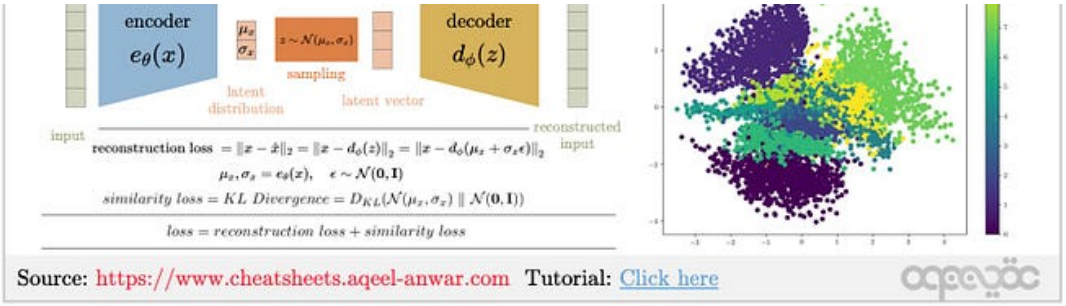




Variational Auto Encoder (VAE)

- Variational autoencoder addresses the issue of non-regularized latent space in autoencoder and provides the generative capability to the entire space.
- Instead of outputting the vectors in the latent space, the encoder of VAE outputs parameters of a pre-defined distribution in the latent space for every input.
- The VAE then imposes a constraint on this latent distribution forcing it to be a normal distribution.
- The latent variable in the compressed form is mean and variance
- The training loss of VAE is defined as the sum of the reconstruction loss and the similarity loss (the KL divergence between the unit gaussian and decoder output distribution.
- The latent variable is smooth and continuous i.e., random values of latent variable generates meaningful output at the decoder, hence the latent space has generative capabilities.
- The input of the decoder is sampled from a gaussian with mean/variance of the output of encoder.





b) Example Questions:

1. What is an Autoencoder?
2. Is the latent space of Autoencoder regularised?
3. What is the loss function for a variational autoencoder?
4. Whats the difference between an Autoencoder and Variational Autoencoder?

c) Detailed Articles:

- Difference between AutoEncoder (AE) and Variational AutoEncoder (VAE)

Summary

This article provides a list of cheat sheets covering important topics for a Machine learning interview followed by some example questions. The list of topics and the number of cheat sheets are constantly being added to the article.

If this article was helpful to you, feel free to clap, share and respond to it. If you want to learn more about Machine Learning and Data Science, follow me @ or connect with me on *LinkedIn*.