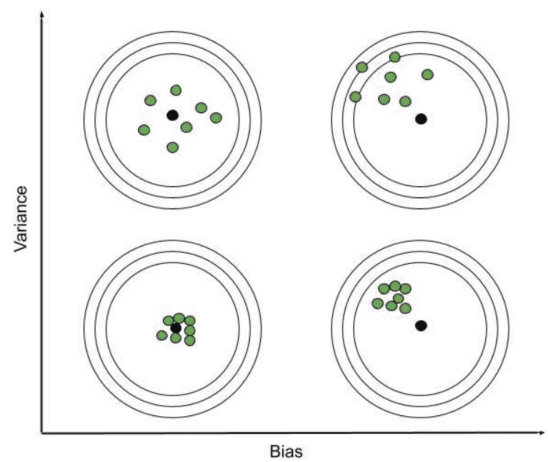
Bias-Variance Tradeoff

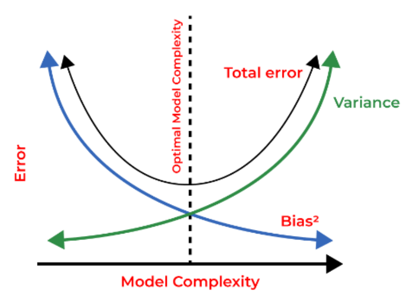
1. Bias & Variance
    a. Basics
        i. **Bias** is a type of error due to wrong assumptions about data; **Variance** reflects model's output sensitivity to variations in training data.
        ii. **Bias** – systematic error due to wrong assumptions
        iii. Both bias and variance are type of *reducible error*
        iv. Importance of understanding Bias and Variance –
        v. Helps in parameter tuning
        vi. Helps in deciding better-fitted models amongst several we trained
    b. **Bias**
        i. What is it –
            1. Difference between expected value and actual value. $\text{Bias}(\hat{Y}) = E(\hat{Y}) - Y$
            2. How far a model's predictions are from the actual values.
            3. **Low Bias** – few assumptions – model is able to capture the underlying patterns
            4. **High Bias** – more assumptions – model is unable to capture underlying patterns - underfitting
        ii. How to reduce Bias –
            1. Use more complex model
            2. Increase number of features
            3. Increase size of training data
            4. Reduce Regularization
    c. **Variance**
        i. What is it –
            1. Amount by which the performance of model changes when it is trained on different subsets of data. $\text{Variance} = E[(\hat{Y} - E[\hat{Y}])^2]$
            2. How much the predictions of a model vary for a given data point when model is trained on different subsets of the same dataset.
            3. **Low Variance** – model is less sensitive to changes in training data
            4. **High Variance** – model is more sensitive to changes - Overfitting
        ii. Ways to reduce it –
            1. CV – use it to tune hyperparameters
            2. Feature selection – reduce num of features
            3. Make model less complex
            4. Regularization
            5. Ensemble
            6. Increase training data
        iii. For common algorithms

1. Linear Regression – High Bias Low Variance
2. Logistic Regression – High Bias Low Variance
3. Decision Trees– Low Bias High Variance
4. Random Forest – Low Bias High Variance
5. Neural Networks Trees– Low Bias High Variance

d. Combinations of Bias-Variance
   i. **Low bias** – means model captures the underlying patterns of data: **in-graph** – points are spread around center point,
   ii. **Low variance** – means model is less sensitive to changes in training data: **in-graph** – points have large distance from each other
   iii. **High Bias Low Variance** - underfit
   iv. **Low Bias High Variance** - overfit
   v. **High Bias High variance** - Worst
   vi. **Low Bias Low Variance** – Best



e. **Bias-Variance Tradeoff**
   i. If model is simple – it would have high bias and low variance; or if model is too complex – it would have low bias and high variance. If the solution is in between these 2 conditions, it would have a tradeoff between bias and



variance.
   ii. **So Bias-Variance trade off is:** If model is too simple, it will not be able to learn the underlying patterns in the data aka high bias and also it will not be sensitive to different subsets of training data aka low variance. On the other

hand, if the model is too complex, it will be able to learn the underlying patterns in the training data aka low bias but it will also be too sensitive to changes in training subset aka high variance. This is bias variance trade off where we sacrifice one for the other. In real, we want to choose an option which is somewhere between these 2 conditions.

   f. **Prioritizing bias over variance -** where we want consistent performance across datasets. Ex: spam identification – it can have higher bias (meaning some spam emails are missed).
   g. **Prioritizing variance over bias –** where we want to capture intricate patterns. Ex – medical diagnostics – missing out critical patterns can have serious consequences

2. Derivation **Total Error = Bias^2 + Variance + Irreducible error**

$$
\begin{aligned}
\mathrm{MSE} &= (Y - \hat{Y})^2 \\
&= (Y - E(\hat{Y}) + E(\hat{Y}) - \hat{Y})^2 \\
&= (Y - E(\hat{Y}))^2 + (E(\hat{Y}) - \hat{Y})^2 + 2(Y - E(\hat{Y}))(E(\hat{Y}) - \hat{Y})
\end{aligned}
$$

Applying the Expectations on both sides.

$$
\begin{aligned}
E[(Y-\hat{Y})^2] &= E[(Y - E(\hat{Y}))^2 + (E(\hat{Y}) - \hat{Y})^2 + 2(Y - E(\hat{Y}))(E(\hat{Y}) - \hat{Y})] \\
&= E[(Y - E(\hat{Y}))^2] + E[(E(\hat{Y}) - \hat{Y})^2] + 2E[(Y - E(\hat{Y}))(E(\hat{Y}) - \hat{Y})]] \\
&= [(Y - E(\hat{Y}))^2] + E[(E(\hat{Y}) - \hat{Y})^2] + 2(Y - E(\hat{Y}))E[(E(\hat{Y}) - \hat{Y})]] \\
&= [(Y - E(\hat{Y}))^2] + E[(E(\hat{Y}) - \hat{Y})^2] + 2(Y - E(\hat{Y}))[E[E(\hat{Y})] - E[\hat{Y}]] \\
&= [(Y - E(\hat{Y}))^2] + E[(E(\hat{Y}) - \hat{Y})^2] + 2(Y - E(\hat{Y}))[E(\hat{Y})] - E[\hat{Y}]] \\
&= [(Y - E(\hat{Y}))^2] + E[(E(\hat{Y}) - \hat{Y})^2] + 2(Y - E(\hat{Y}))[0] \\
&= [(Y - E(\hat{Y}))^2] + E[(E(\hat{Y}) - \hat{Y})^2] + 0 \\
&= [\mathrm{Bias}^2] + \mathrm{Variance}
\end{aligned}
$$

Irreducible error – ex: noise in measurements of data

3. Bias Variance tradeoff in L1 and L2 regularization
   a. L1 regularization – induces sparsity, drive some coeffs to 0 -> make model simple -> high bias and lower variance
   b. L2 regularization – shrinks coeffs not remove them-> as compared to L1, L2 has lower bias and higher variance
4. How do you measure Variance and Bias in Neural Networks
   a. Use `from mlxtend.evaluate import bias_variance_decomp`
5. How to measure Variance in Linear Regression
   a. Use `from mlxtend.evaluate import bias_variance_decomp`
6. How to measure Bias in Linear Regression
   a. Use `from mlxtend.evaluate import bias_variance_decomp`
7. How to measure bias-variance tradeoff?
8. How to visualize bias-variance tradeoffs?
9. How to reduce bias-variance tradeoffs?
10. How to demonstrate bias-variance tradeoffs?
11. What are sources of error?

    a. Assumptions (bias) + model choice (variance) + noise (irreducible)

12. Why is it impossible to minimize both bias and variance simultaneously?
    a. Already answered