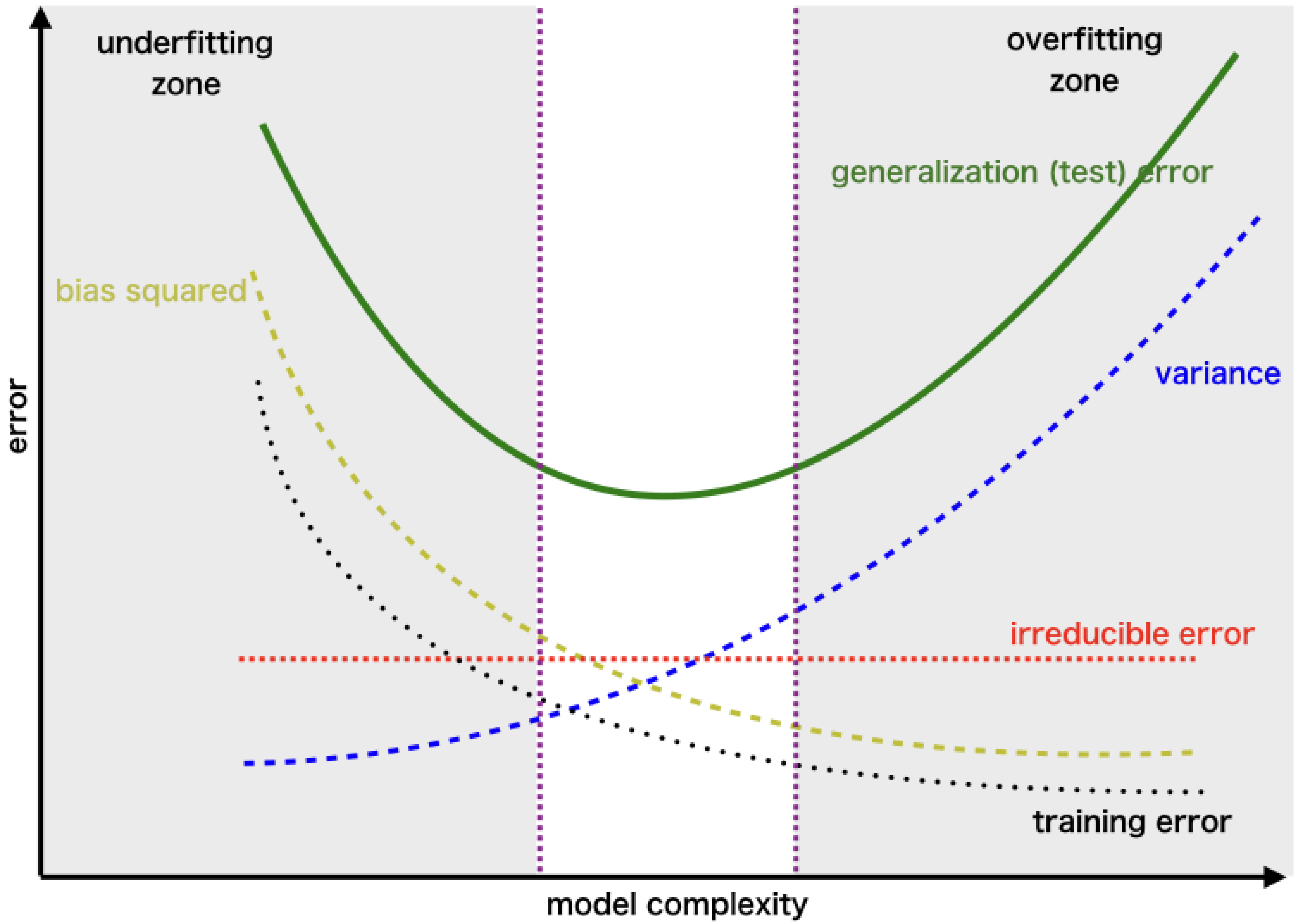
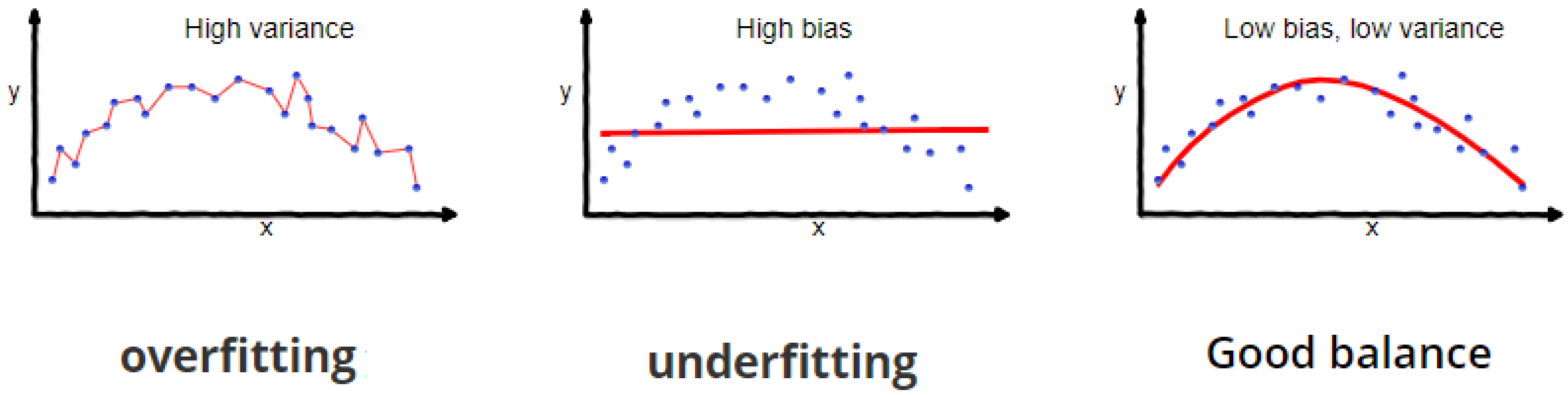
In [5]:
```
# Bias Variance and tradeoff
# Prediction Error has 3 components - Bias, Variance and Irreducible error.

# Irreducible Error cannot be reduced regardless of which algorithm is used. This error
# originates from wrong framing of business problems and also caused by unknown variables

# Reducible Error is that error which can be minimized by improving model itself like
# adjusting parameters in the algorithm. Model optimization is key. Bias and variance are
# associated with this error

# Bias occurs due to wrong assumptions about data such as assuming data is linear whereas
# in reality data is complex and follows non linear distribution. Bias are simplifying
# assumptions made by an algorithm to predict and makes it easier for cost function to
# learn effectively.

# Bias results in consistent errors and can cause a model to be too simple and miss key
# relationships in data
# Bias leads to Underfitting

# Linear Algorithms have too much bias like Multiple linear regression assumes normal
# distribution and predictive performance will be bad if assumptions are not met.

# Low bias fewer assumptions and high bias many assumptions.
# High Bias Algorithms are Multiple Linear Regression, Binary Logistic Regression
# Low Bias Algorithms are Tree based models like Decision Tree, Random Forest

# Variance is deviation caused in model when there are minor to major changes in input
# variables or independent variables. Model is too sensitive to small fluctuations in
# training data. Results in incorrect predictions

# Model becomes too complex and captures noise instead of underlying patterns in data.
# Variance leads to Over fitting
# Low variance means model is less sensitive to changes in training data and produces
# consistent predictions
# High variance means model is very sensitive to changes in training data and results
# in significant changes in predictions

# Underfitting - High Bias Low Variance
# Overfitting - Low Bias High Variance
```



In [9]:
```
# Underfitting is when model fails to capture the underlying data pattern and results in
# poor performance both on train and test data.
# Metrics like R Square (<0.60) and Accuracy (<0.70) that are below minimum threshold
# indicate underfitting.

# Occurs due to insufficient data, inadequate traing time, simple model, poor feature
```

```
# engineering or variable selection and excessive regularization

# Ways to reduce uderfitting and also bias are
# a) Use a complex model that captures complexity in data like using ensemble methods,
# neural networks rather than  base models.
# b) Increase Number of features or variables - Add mode variables to training data that will
# improve algorithm ability to capture complex patterns
# c) Increase size of training data - Increasing number of observations reduces bias
# as algorithm has more data to train
# d) Reduce regularization using regularization methods like L1 (LASSO) & L2(Ridge) help
# improve model performance.

# Overfitting occurs when a machine learning model fits training data too closely and making
# it unable to generalize on new data. Model performs well on training data but performs
# poorly on test data or new data.
# Typically RSquare(>0.95) and Accuracy(>0.95) close to 1 indicates overfitting

# Ways to reduce overfitting and variance are
# a) Cross Vaildation - Helps in identifying whether a model overfitting or not. K fold
# Cross validation where in random sub samples of data created from different parts of
# training data and these sub samples are used for repeated testing of model.

# b) Proper feature selection by choosing only relevant variables

# c) Use regularization like L1 & L2

# d) Ensemble Methods like Boosting and Bagging

# e) Simplifying model by reducing complexity

# f) Early Stopping is  stopping training model once a threshold performance is reached
```



In [13]:
```
# Regularization is atechnique used to control overfitting. Regularization adds a penalty
# to model for being too complex, encouraging it to stay simple and more genral

# Regulariuzation adds as penalty term called as lambda or alpha to regression object or
# coefficeints of the regression model

# penalization will either shrink size of coefficients or make some coefficients equal to
# Zero.
```

In [17]:
```
# L1 Regularization or LASSO - Least Absolute Shrinkage Selection Operator
# LASSO adds a penalty to absolute value of magnitude of coefficients of regression model
# LASSO will definitely make some coefficients equal to Zero and hence can be used as
# feature or variable selection method
# LASSO doesnot work well with multicolliear data
# lambda or Alpha must be between 0 to infinity. If Zero none coefficients eliminated
# and if infinity all coefficients eleiminated
```

$$Cost\ (W) = RSS(W) + \lambda * (sum\ of\ absolute\ value\ of\ weights)$$

$$= \sum_{i=1}^{N} \left\{ y_i - \sum_{j=0}^{M} w_j\, x_{ij} \right\}^2 + \lambda \sum_{j=0}^{M} |w_j|$$

In [20]:
```
# L2 or Ridge Regression adds a penalty lamba/alpha to sum square or squared magnitude of
# coefficients.
# Ridge regression shrinks size of coefficients but doesnot make them Zero and cannot
# be used as variable/feature selection method.
# Ridge regression works well with multicollinear data
```

$$Cost\ (W) = RSS(W) + \lambda * (sum\ of\ squares\ of\ weights)$$

$$= \sum_{i=1}^{N} \left\{ y_i - \sum_{j=0}^{M} w_j\, x_{ij} \right\}^2 + \lambda \sum_{j=0}^{M} w_j^2$$

In [24]:
```
# Elastic Net is combination of L1 & L2 regularization .
# First LASSO is run and on its output Ridge Regression is done.
```

In [ ]:
```
```