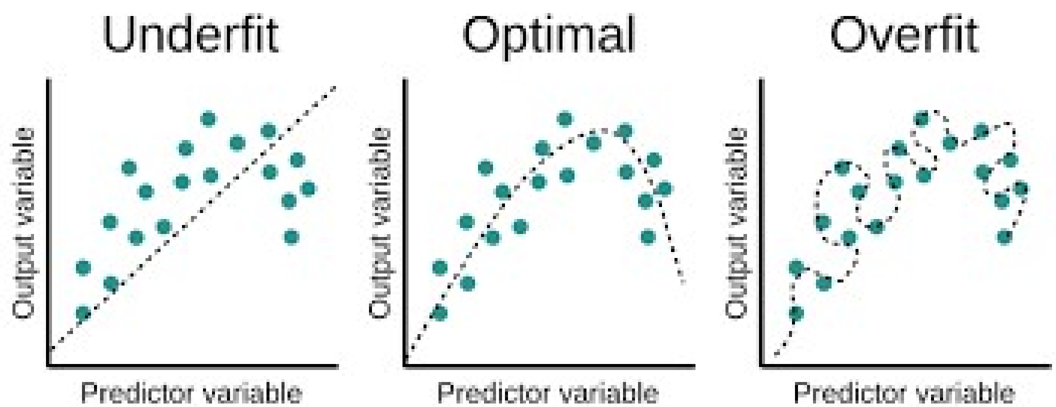
# Regularization

# Regularization



# Regularization

- Regularization is a technique used in machine learning to prevent overfitting and improve the generalization performance of models.
- Overfitting occurs when a model learns to fit the training data too closely, capturing noise or random fluctuations that are specific to the training data and do not generalize well to unseen data.
- Regularization helps address this issue by adding a penalty term to the model's objective function, encouraging simpler models and reducing the risk of overfitting.

# Regularization

- Regularization works by adding a penalty term to the model's loss function, which discourages the model from learning overly complex patterns.
- This penalty is typically based on the model's parameters.
- There are two common types of regularization used in regression:

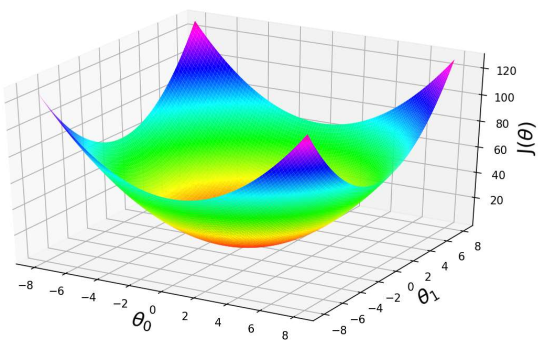
# LASSO Regression

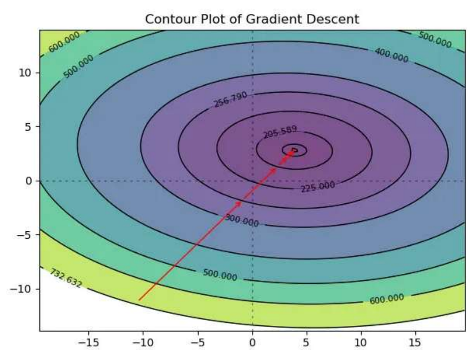
- Least Absolute Shrinkage and Selection Operator – L1 Regularization.

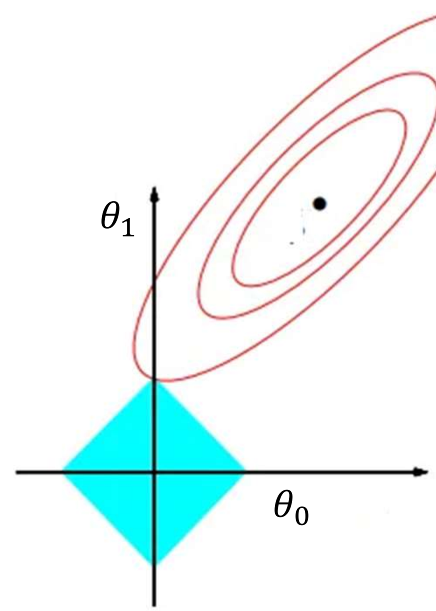
# LASSO Regression

- The cost function is
- $J(\bar{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(\sum_{j=0}^{d}\theta_j\, x_j^i - y^i\right)^2 + \lambda\left|\bar{\theta}\right|_1$
- $\frac{\partial J(\bar{\theta})}{\partial \theta_j} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y}^i - y^i) * x_j^i + \lambda * sign(\bar{\theta})$

$\lambda(|\theta_0| + |\theta_1| + \cdots + |\theta_d|)$



# LASSO Regression

- This penalty term tends to shrink parameters towards zero, but importantly, it also has the property of setting some parameters exactly to zero when the regularization parameter $\lambda$ is sufficiently large.
- This property of Lasso regularization makes it useful for feature selection, as it effectively removes irrelevant features by setting their corresponding coefficients to zero.

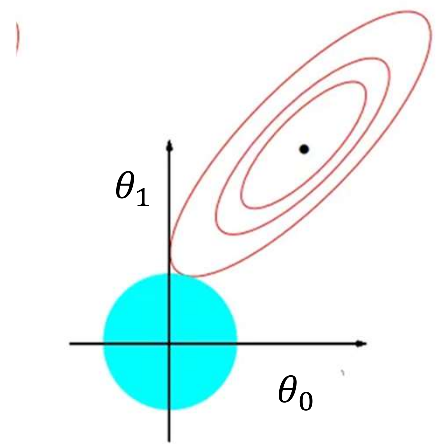
# Ridge Regression

- Ridge Regression – L2 Regularization
- Ridge regression, also known as Tikhonov regularization, is a linear regression technique that adds a penalty term to the $MSE$ objective function.
- This penalty term is based on the L2 norm (Euclidean norm) of the parameter vector.

# Ridge  Regression

- The cost function is
- $J(\bar{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(\sum_{j=0}^{d} \theta_j \, x_j^i - y^i\right)^2 + \lambda||\bar{\theta}||_2$
- $\frac{\partial J(\bar{\theta})}{\partial \theta_j} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y^i} - y^i) * x_j^i + 2 * \lambda * \bar{\theta}$

$$\lambda\sqrt{\theta_1^2 + \theta_2^2 + \cdots + \theta_d^2}$$



# Ridge Regression

- Ridge regression uses L2 regularization, which penalizes the sum of the squared parameters, encouraging them to be small but not exactly zero.
- The effect of Ridge regularization is to shrink the coefficients towards zero, with larger coefficients being shrunk more.
- This helps to reduce the variance of the model and can improve its generalization performance, especially when dealing with multicollinearity (high correlation among predictor variables).