

- Boxplot:
 - $Q_1 = N \times 0.25$
 - $Q_2 = N \times 0.5$
 - $Q_3 = N \times 0.75$
 - $IQR = Q_3 - Q_1$
 - $Bounds = [Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$
- Pearson = $\frac{\Sigma(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sqrt{\Sigma(y_{1i} - \bar{y}_1)^2 \times \Sigma(y_{2i} - \bar{y}_2)^2}}$
- Spearman: Assign ranks and apply Pearson formula.
Example: $[20, 10, 20, 30, 20] \rightarrow [3, 1, 3, 5, 3]$
- Information Gain: $IG(y_{out}|y_i) = E(y_{out}) - E(y_{out}|y_i)$
- Entropy: $E(y) = -\sum P(x_i) \log(P(x_i))$
- Normalization:
 - MinMax: $\frac{y_i - \min}{\max - \min}$
 - Standardization: $\frac{y_i - \mu}{\sigma}$
- Binarization:
 - Range (equal width): Depends on variable range
Example: $y \in [-1, 1] : [0.2, -0.1, 0.6] \rightarrow [1, 0, 1]$
 - Frequency (equal depth): Depends on variable mean
Example: $\bar{y} = 25 : [10, 40, 30, 20] \rightarrow [0, 1, 1, 0]$
- Confusion Matrix:

		True		
		A	B	C
Pred	A	TA	FA	FA
	B	FP	TB	FB
	C	FC	FC	TC
- Metrics:
 - Accuracy = $\frac{TP + TN}{total}$
 - Error rate = $1 - Accuracy = \frac{FP + FN}{total}$
 - Recall = $\frac{TP}{TP + FN}$ (Sensitivity)
 - Fallout = $\frac{TN}{N} = \frac{TN}{TN + FP}$ (Specificity)
 - Precision = $\frac{TP}{TP + FP}$
 - $F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$
- Error:
 - MSE = $\sum (Z - \hat{Z})^2$
 - RMSE = \sqrt{MSE}
 - MAE = $\sum |Z - \hat{Z}|$
- Decision trees:
 1. Choose feature with highest IG.
 2. Split dataset by that feature, create leaves if necessary.
 3. Repeat until unable to proceed.

Prune: (Given a twig)

 1. Count it's leaves labels.
Example: $\#A = 5, \#B = 6$
 2. Remove it's leaves.
 3. Relabel twig as a leaf.
Example: $B(6/11), \#B > \#A$
- Gaussian Distribution:
 - Mean: $\mu = \frac{1}{n} \sum y_i$
 - Standard Deviation: $\sigma = \sqrt{\frac{\sum (y_i - \mu)^2}{n - 1}}$
 - $P(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right)$

- Naive Bayes:
 - MAP: $P(C|x) = \frac{P(C)P(x|C)}{P(x)}$
 - ML: $P(C|x) = P(x|C)$
 - 1. Calculate probability for each class.
 - 2. Calculate $P(x|C)$ for each attribute and each class.
 - 3. TODO
- K-Nearest Neighbors:
 1. Distances: (for n variables)
 - Manhattan: $\sum |y_{1i} - y_{2i}|$
 - Euclidean: $\sqrt{\sum (y_{1i} - y_{2i})^2}$
 - Cosine: $\frac{\sum y_{1i} y_{2i}}{\sqrt{\sum y_{1i}^2} \sqrt{\sum y_{2i}^2}}$
 - Hamming: #Differences
 2. Choose K nearest neighbors.
 3. Classify using mean if variable is numeric, or mode if it is categorical.
 4. If weighted, divide by weight.
- Regressions:
 - Linear: $W = (X^T X)^{-1} X^T Z$
 - Ridge: $W = (X^T X + \lambda I)^{-1} X^T Z$
- Perceptron:

$\hat{Z} = a(W^T X)$, $a \leftarrow$ activation function
Se $Z \neq \hat{Z} \rightarrow W' = W + \eta(Z - \hat{Z})X$
- Neural Networks (MLP):
 - Forward: $x^{[0]} \rightarrow z^{[1]} = w^{[1]}x^{[0]} + b^{[1]} \rightarrow x^{[1]} = a(z^{[1]}) \rightarrow \dots \rightarrow z^{[i]} = w^{[i]}x^{[i-1]} + b^{[i]} \rightarrow x^{[i]} = a(z^{[i]}) \rightarrow E$
 - Backward:
 - * $\delta^{[last]} = \frac{\partial E}{\partial x^{[last]}} \circ \frac{\partial x^{[last]}}{\partial z^{[last]}}$
 - * $\delta^{[i]} = \left(\frac{\partial z^{[i+1]}}{\partial x^{[i]}}\right)^T \cdot \delta^{[i+1]} \circ \frac{\partial x^{[i]}}{\partial z^{[i]}}$
 - * $w^{[i]'} = w^{[i]} - \eta \frac{\partial E}{\partial w^{[i]}}$
 - * $\frac{\partial E}{\partial w^{[i]}} = \delta^{[i]} \cdot \left(\frac{\partial z^{[i]}}{\partial w^{[i]}}\right)^T$
 - * $b^{[i]'} = b^{[i]} - \eta \frac{\partial E}{\partial b^{[i]}}$
 - * $\frac{\partial E}{\partial b^{[i]}} = \delta^{[i]}$
 - * $\frac{\partial z^{[i+1]}}{\partial x^{[i]}} = w^{[i+1]} \quad \frac{\partial z^{[i]}}{\partial w^{[i]}} = x^{[i-1]} \quad \frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$

Name	Error function	$\frac{\partial E}{\partial x^{[i]}}$
Squared Error	$\frac{1}{2} \left(x^{[i]} - t\right)^2$	$x^{[i]} - t$
Cross-entropy	$-\sum_{i=1}^n t_i \log\left(x_i^{[i]}\right)$	$x^{[i]} - t$
Name	Activation function	$\frac{\partial x^{[i]}}{\partial z^{[i]}}$
Sigmoid	$\frac{1}{1 + e^{-x}}$	$\sigma(z^{[i]})(1 - \sigma(z^{[i]}))$
Hyperbolic tangent	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	$1 - \tanh\left(z^{[i]}\right)^2$

- Gaussian Mixture:

- Covariance Matrix: $\Sigma = \begin{bmatrix} TODO & TODO \\ symmetric & TODO \end{bmatrix}$

- $P(y|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \cdot \exp\left(-\frac{1}{2}(y - \mu)^T \Sigma^{-1}(y - \mu)\right)$

- K-Means:

TODO

- EM:

TODO