- 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 $= \dfrac{\Sigma(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sqrt{\Sigma(y_{1i} - \bar{y}_1)^2 \times (y_{2i} - \bar{y}_2)^2}}$
- Spearman: Assign ranks and apply Pearson formula.
  Example: [20, 10, 20, 30, 20] → [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: $\dfrac{y_i - min}{max - min}$
  - Standardization: $\dfrac{y_i - \mu}{\sigma}$
- Binarization:
  - Range (equal width): Depends on variable range
    Example: $y \in [-1, 1] : [0.2, -0.1, 0.6] \to [1, 0, 1]$
  - Frequency (equal depth): Depends on variable mean
    Example: $\bar{y} = 25 : [10, 40, 30, 20] \to [0, 1, 1, 0]$
- Confusion Matrix:

		True		
		A	B	C
Pred	A	TA	FA	FA
	B	FP	TB	FB
	C	FC	FC	TC

- Metrics:
  - Accuracy $= \dfrac{TP + TN}{total}$
  - Error rate $= 1 - Accuracy = \dfrac{FP + FN}{total}$
  - Recall $= \dfrac{TP}{TP + FN}$ (Sensitivity)
  - Fallout $= \dfrac{TN}{N} = \dfrac{TN}{TN + FP}$ (Specificity)
  - Precision $= \dfrac{TP}{TP + FP}$
  - $\text{F}_1 = \dfrac{TP}{TP + \frac{1}{2}(FP + FN)}$
- 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: $B(6/11),\quad \#B > \#A$
- Gaussian Distribution:
  - Mean: $\mu = \dfrac{1}{n}\sum y_i$
  - Standard Deviation: $\sigma = \sqrt{\dfrac{\sum(y_i - \mu)^2}{n - 1}}$
  - $P(y|\mu, \sigma) = \dfrac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\dfrac{1}{2}\left(\dfrac{y - \mu}{\sigma}\right)^2\right)$
- Bayes Rule:
  - MAP: $P(C|x) = \dfrac{P(C)P(x|C)}{P(x)}$
  - ML: $P(C|x) = P(x|C)$

  1. TODO
- K-Nearest Neighbors:
  1. Distances: (for n variables)
     - Manhattan: $\sum|y_{1i} - y_{2i}|$
     - Euclidean: $\sqrt{\sum(y_{1i} - y_{2i})^2}$
     - Cosine: $\dfrac{\sum y_{1i}\ y_{2i}}{\sqrt{\sum y_{1i}^2}\sqrt{\sum y_{2i}^2}}$
     - Hamming: #Differences
  2. Choose K nearest neighbors.
  3. Classify using mean if variable is numeric, or mode if it is categoric.
  4. If weighted, divide by weight.
- Regressions:
  - Linear: $W = (X^TX)^{-1}X^TZ$
  - Ridge: $W = (X^TX + \lambda I)^{-1}X^TZ$
- Perceptron:
  $\hat{Z} = a(W^TX),\ \ a \leftarrow$ activation function
  Se $Z \neq \hat{Z} \longrightarrow W' = W + \eta(Z - \hat{Z})X$
- Neural Networks (MLP):
  - Forward: $x^{[0]} \to z^{[1]} = w^{[1]}x^{[0]} + b^{[1]} \to x^{[1]} = a\left(z^{[1]}\right) \to \ldots \to z^{[i]} = w^{[i]}x^{[i-1]} + b^{[i]} \to x^{[i]} = a\left(z^{[i]}\right) \to E$
  - Backward:
    * $\delta^{[last]} = \dfrac{\partial E}{\partial x^{[last]}} \circ \dfrac{\partial x^{[last]}}{\partial z^{[last]}}$
    * $\delta^{[i]} = \left(\dfrac{\partial z^{[i+1]}}{\partial x^{[i]}}\right)^T \cdot \delta^{[i+1]} \circ \dfrac{\partial x^{[i]}}{\partial z^{[i]}}$
    * $w^{[i]'} = w^{[i]} - \eta\dfrac{\partial E}{\partial w^{[i]}}$
    * $\dfrac{\partial E}{\partial w^{[i]}} = \delta^{[i]} \cdot \left(\dfrac{\partial z^{[i]}}{\partial w^{[i]}}\right)^T$
    * $b^{[i]'} = b^{[i]} - \eta\dfrac{\partial E}{\partial b^{[i]}}$
    * $\dfrac{\partial E}{\partial b^{[i]}} = \delta^{[i]}$
    * $\dfrac{\partial z^{[i+1]}}{\partial x^{[i]}} = w^{[i+1]} \qquad \dfrac{\partial z^{[i]}}{\partial w^{[i]}} = x^{[i-1]} \qquad \dfrac{\partial z^{[i]}}{\partial b^{[i]}} = 1$

Name	Error function	$\dfrac{\partial E}{\partial x^{[i]}}$
Squared Error	$\dfrac{1}{2}\left(x^{[i]} - t\right)^2$	$x^{[i]} - t$
Cross-entropy	$-\sum_{i=1}^{n} t_i \log\left(x_i^{[i]}\right)$	$x^{[i]} - t$

Name	Activation function	$\dfrac{\partial x^{[i]}}{\partial z^{[i]}}$
Sigmoid	$\dfrac{1}{1 + e^{-x}}$	$\sigma(z^{[i]})(1 - \sigma(z^{[i]}))$
Hyperbolic tangent	$\dfrac{e^x - e^{-x}}{e^x + e^{-x}}$	$1 - \tanh\left(z^{[i]}\right)^2$

- Gaussian Mixture:
  - Covariance Matrix: $\Sigma = \begin{bmatrix} TODO & TODO \\ symmetric & TODO \end{bmatrix}$
  - $P(y|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \cdot exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(u-\mu)\right)$
- K-Means:
  TODO
- EM:
  TODO