Classification Accuracy $= \frac{N_{correct}}{N_{total}}$

Classification Error $= \frac{N_{miss}}{N_{total}}$

Precision = PPV $= \frac{TP}{TP+FP}$

Recall = TPR $= \frac{TP}{TP+FN}$

Fallout = FPR = $\frac{FP}{FP+TN}$

Specificity = TNR = $\frac{TP}{TN+FP}$

F1-score = $2 \times \frac{precision \times recall}{precision+recall}$

ROC: TPR y-axis, FPR x axis

AUC: P(error)$= P(C_i) \cdot FP \; + \; P(C_i) \cdot FN$

FP/FN = $\frac{\delta - mean}{\sqrt{varience}}$

FLDA: $J(w) = \frac{(m_1 - m_2)^2}{\varsigma_1^2 + \varsigma_2^2} \rightarrow \frac{w^T S_B w}{w^T S_W w}$

$S_B = (m_1 - m_2)^2, S_W \; = \sigma_1 + \sigma_2$ (within class variance)

$$W_0 = \frac{-(m_1 + m_2)^T W}{2}$$

Optimized:$S_W^{-1} S_B w = \lambda w$, eigenvalues

Max projection: N -1 classes

Perceptron Error: $E_p(w,b) = -\sum_{n \varepsilon M} (w^T x_n + b) t_n$

Stochastic Gradient Decent:

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \frac{\partial E_p(w,b)}{\partial w} = w^{(t)} + \eta x_n t_n$$

$$b^{(t+1)} \leftarrow b^{(t)} - \eta \frac{\partial E_p(w,b)}{\partial b} = b^{(t)} + \eta t_n$$

Where η is the learning rate

Logistic Regression: $\phi(x) = \frac{1}{1+\exp(-x)}$

Objective function: (Binary) Cross Entropy:

$$J(w, w_0) = \Sigma_{i=1}^{N} - t_i \log \phi\,(z_i) \\ - (1 - t_i) \log\big(1 - \phi(z_i)\big)$$

$$w^{(t+1)} \leftarrow w^{(t)} + \eta \sum_{i=1}^{N} (t_i - y_i) x_i$$

$$w_0{}^{(t+1)} \leftarrow w_0^{(t)} + \eta \sum_{i=1}^{N} (t_i - y_i)$$

SoftMax: $p_k = \frac{\exp(y_k(x))}{\Sigma_{j=1}^{K} \exp(y_j(x))}$

where $y_k(x) = \phi(w_k^T \, x + b_k)$

Hard Margin SVM: $y(x) = w^T \phi(x) + b$

$\arg_{w,b} \min \frac{1}{2} ||w||^2, t_n y(x_n) \geq 1$

Polynomial Kernel:$K(x,y) = (1 + \langle x, y \rangle)^d$

RBF kernel: $K(x,y) = e^{-\gamma ||x-y||^2}$

Sigmoid kernel: $K(x,y) = \tanh(\gamma(x,y) + 1)$

Kernal Trick: $y(x) = \sum_{n=1}^{N} a_n \, t_n K(x, x_n) + b$

Constraints: $a_n \geq 0$

$t_n y(x_n) \geq 0$

$a_n(t_n y(x_n) - 1) = 0$

Threshold:

$$b = \frac{1}{N_S} \Sigma_{n \varepsilon S} \big(t_n - \Sigma_{m \varepsilon S} a_m t_m K(x_n, x_m)\big)$$

Lagrange Multiplier: $L(w,b,a) \frac{1}{2} ||w||^2 - \sum_{n=1}^{N} a_n (t_n (w^T \phi(x_n) + b) - 1)$

Conditions: $w = \sum_{n=1}^{N} a_n \, t_n \phi(x_n)$

$0 = \sum_{n=1}^{N} a_n \, t_n$

Soft Margin SVM: $\arg_{w,b} \min \quad \frac{1}{2} ||w||^2 + C \sum_{n=1}^{N} \xi_n$

Slack Variable (ξ_n): $t_n y(x_n) \geq 1 - \xi_n,$ on the wrong side then > 1, correct side or on the line = 0, inside the correct margin < 1.

Dual Lagrangian: $L(a) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n \, a_m t_n t_m K(x_n, x_m) \quad b = w_0 = \frac{1}{N_M} \sum_{n \in M} \big(t_n - \sum_{m \in S} a_m \, t_m K(x_n, x_m)\big)$

PCA: Max var: $a_1^T R_x a_1 = \lambda_1$

Min error: $\sum_{j=m+1}^{D} a_j^T R_x a_j \rightarrow R_x a_j = \lambda_j a_j$

$\sum_{j=m+1}^{D} a_j^T R_x a_j \rightarrow R_x a_j = \lambda_j a_j$, choose the largest eigenvalue. Assume Gaussian distribution and mean centered.

Backpropagation: $\frac{\partial J}{\partial W} = \frac{\partial V}{\partial W} \cdot \frac{\partial y}{\partial V} \cdot \frac{\partial E}{\partial y} \cdot \frac{\partial J}{\partial E}$

Chain Rule:

$$J(w) = \frac{1}{2}\sum_{i=1}^{N} e_i^2 \rightarrow e_i(-1)\phi'(v_i)x_j$$

$$e_i = t_i - y_i$$

$y_i = \phi(v_i)$, ϕ is an activation func.

$v_i = w^T x_j$, note that $x_j \varepsilon R^{D+1}$

$$w_{lj}^{(t+1)} = w_{lj}^{(t)} - \eta \frac{\partial J}{\partial w_{lj}}$$
$$= w_{lj}^{(t)} + \eta e_i \phi'(v_l) x_j$$

Momentum: $w^{(t+1)} = w^{(t)} + \Delta w^{(t)}$ where

$\Delta w^{(t)} = -\eta \nabla J\left(w^{(t)}\right) + \alpha \Delta w^{(t-1)}$, $\alpha = 0.9$

Nesterov's momentum: $w^{(t+1)} = w^{(t)} + \Delta w^{(t)}$ where

$\Delta w^{(t)} = -\eta \nabla J\left(m^{(t)}\right)$, $m^{(t)} = w^{(t)} + \alpha \Delta w^{(t-1)}$

Activation function	$\phi(x)$	$\phi'(x) = \frac{d\phi(x)}{dx}$
Linear	x	1
Sigmoid/Logistic	$\frac{1}{1+e^{-x}}$	$\phi(x)(1-\phi(x))$
Tanh	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	$1-\phi(x)^2$
ReLU	$\begin{cases} 0, & x \le 0 \\ x, & x > 0 \end{cases}$	$\begin{cases} 0, & x < 0 \\ 1, & x > 0 \\ \text{undefined}, & x = 0 \end{cases}$
Leaky ReLU	$\begin{cases} 0.01x, & x \le 0 \\ x, & x > 0 \end{cases}$	$\begin{cases} 0.01, & x < 0 \\ 1, & x > 0 \\ \text{undefined}, & x = 0 \end{cases}$
Softplus	$(1+e^x)$	$\frac{1}{1+e^{-x}}$
ELU	$\begin{cases} \alpha(e^x - 1), & x \le 0 \\ x, & x > 0 \end{cases}$	$\begin{cases} \alpha e^x, & x < 0 \\ 1, & x > 0 \\ 1, & x > 0 \text{ and } \alpha = 1 \end{cases}$
SELU	$\lambda \begin{cases} \alpha(e^x - 1), & x < 0 \\ x, & x \ge 0 \end{cases}$	$\lambda \begin{cases} \alpha e^x, & x < 0 \\ 1, & x \ge 0 \end{cases}$