$$Loss(\hat{y}_i, y_i) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \cdot \log(\hat{y}_i) + (1-y_i)\cdot \log(1-\hat{y}_i)\right]$$

$$\hat{y} = \frac{e^{(b_0 + X \cdot W)}}{1 + e^{(b_0 + X\cdot W)}}$$

$$\begin{aligned} b_0 + X \cdot W \rightarrow & \theta^T \cdot X \\ = & \theta_0 \cdot x_0 + \theta_1 \cdot x_1 + \cdots + \theta_n \cdot x_n \end{aligned}$$

$$\rightarrow \hat{y} = \frac{e^{\theta^T \cdot X}}{1 + e^{\theta^T \cdot X}} \overset{\times \frac{\frac{1}{e^{\theta^T \cdot X}}}{\frac{1}{e^{\theta^T \cdot X}}}}{=} \frac{1}{1 + e^{-\theta^T \cdot X}}$$

$$\hat{y} = \frac{1}{1 + e^{-\theta^T \cdot X}}$$

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \cdot \log(\hat{y}_i(\theta)) + (1-y_i)\cdot \log(1-\hat{y}_i(\theta))\right]$$

$$\frac{\partial}{\partial \theta_j} J(\vec{\theta})$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right) x_{i,j}$$

$$\theta_j \rightarrow \theta_j - \eta \cdot \frac{\partial}{\partial \theta_j} J(\vec{\theta})$$

$$f = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$$

$$\text{Log loss} = -y_i \cdot \log(\hat{y}_i) - (1-y_i)\cdot \log(1-\hat{y}_i)$$

$$-\sum_{s=1}^{C=2} y_{s,i} \cdot \log(\hat{y}_{s,i}) = -y_{1,i} \cdot \log(\hat{y}_{1,i}) - y_{2,i} \cdot \log(\hat{y}_{2,i})$$

$$y_{1,i} = (1 - y_{2,i})$$

$$\hat{y}_{1,i} = (1 - \hat{y}_{2,i})$$

$$\sum_{s=1}^{C=2} y_{s,i} \cdot \log(\hat{y}_{s,i}) \rightarrow \sum_{s=1}^{C=N} y_{s,i} \cdot \log(\hat{y}_{s,i})$$

$$extra = \lambda \sum_{i=1}^{n} \theta_i^2$$

$$extra = \frac{\lambda}{2} \sum_{i=1}^{n} \theta_i^2$$

$$extra = \lambda \sum_{i=1}^{n} |\theta_i|$$

$$extra = r\lambda \sum_{i=1}^{n} |\theta_i| + (1-r) \cdot \frac{\lambda}{2} \sum_{i=1}^{n} \theta_i^2$$

$$\frac{\partial}{\partial \theta} \rightarrow \lambda \sum_{i=1}^{n} \theta_i$$

$$loss = \max(0, 1 - t_i \cdot \hat{y}_i);\; t_i = \pm 1$$

$$\hat{y} = 0 \text{ if } \vec{w} \cdot \vec{x} + b < 0$$
$$\hat{y} = 1 \text{ if } \vec{w} \cdot \vec{x} + b \geq 0$$

$$\textbf{Minimze } \frac{1}{2}\vec{w}^2 + C \sum_{i=1}^{m} \zeta_i; \quad \text{for } \vec{w}, b, \zeta$$
$$\textbf{constrained by } t_i(\vec{w} \cdot x_i + b) \geq 1 - \zeta_i$$

$$K(x, x') = x \cdot x'$$

$$K(x, x') = (\gamma x \cdot x' + c)^d$$

$$K(x, x') = exp(-\gamma ||x - x'||^2)$$

$$K(x, x') = \tanh(\gamma x \cdot x' + c)$$

$$d_m = |x_1 - x_2| + |y_1 - y_2|$$

$$d_e = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$d_L = \sqrt[m]{(x_1 - x_2)^m + (y_1 - y_2)^m}$$
$$d_L = (|x_1 - x_2|^m + |y_1 - y_2|^m)^{\frac{1}{m}}$$

$$G = 1 - \sum_{i=1}^{K} p_i^2$$

$$p_i$$

$$H = -\sum_{i=1}^{K} p_i \log_2 p_i$$

$$J = \frac{n_{left}}{N} \cdot G_{left} + \frac{n_{right}}{N} \cdot G_{right}$$

$$R_\alpha(T) = R(T) + \alpha|T|$$

$$P(y|x_1, ...x_n) = \frac{P(x_1, ..., x_n|y)P(y)}{P(x_1, ..., x_n)}$$

$$P(x_i|y, x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) = P(x_i|y)$$

$$a_i$$

$$P(y|x_1, ...x_n) = P(y)\frac{\prod_{i=1}^{n} P(x_i|y)}{P(x_1, ..., x_n)}$$

$$p(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

$$P(x_i|y) = \frac{x_i}{N}$$
$$\rightarrow P(x_i|y) = \frac{x_i + \alpha}{N + \alpha \cdot n}$$

$$P(x_i|y) = \frac{N_{x_i,y}}{N_y}$$
$$\rightarrow P(x_i|y) = \frac{N_{x_i,y} + \alpha}{N_y + \alpha \cdot n}$$

$$\begin{aligned}
\hat{y} &= \theta_0 + \theta_1 \cdot x_1 + \cdots + \theta_m \cdot x_m + \vec{\epsilon} \\
\hat{y} &= \vec{X}^T \cdot \vec{\theta} + \vec{\epsilon} \\
\vec{\theta} &= \text{ weights} \\
\hat{y} &= \text{ prediction} \\
\vec{X} &= \text{ features} \\
\vec{\epsilon} &= \text{ error}
\end{aligned}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\begin{aligned}
P(A) &- \text{ probability of A} \\
P(B) &- \text{ probability of B} \\
P(A \cap B) &- \text{ probability of A \& B} \\
P(A|B) &- \text{ probability of A given B}
\end{aligned}$$

$$\begin{aligned}
P(F) &= \frac{63 + 18}{96} = \frac{81}{96} \\
P(NF) &= \frac{15 + 18}{96} = \frac{33}{96} \\
P(F \cap NF) &= \frac{18}{96}
\end{aligned}$$

$$\begin{aligned}
P(F|NF) &= \frac{P(F \cap NF)}{P(NF)} = \frac{18}{33} \\
P(NF|F) &= \frac{P(F \cap NF)}{P(F)} = \frac{18}{81}
\end{aligned}$$

$$P(A|B) = \frac{P(B|A)}{P(B)} \cdot P(A)$$

$$\begin{aligned}
P(A) - &\ \text{probability of A} \\
P(B) - &\ \text{probability of B} \\
P(B|A) - &\ \text{probability of B given A} \\
P(A|B) - &\ \text{probability of A given B}
\end{aligned}$$

$$P(C|+) = \frac{P(+|C)}{P(+)} \cdot P(C) = \frac{0.995}{0.03193} \cdot 0.002 = 0.062$$

$$\begin{aligned}
\beta :&\ \text{false negative rate} \\
\text{power } &= 1 - \beta \\
\text{power } &= Pr(\text{reject } H_0 \mid H_1 \text{ true})
\end{aligned}$$

$$\begin{aligned}
y_{i,t} &\quad \text{i = class, t = time} \\
\bar{y}_{i,t} &\quad \text{mean of } y_{i,t} \\
\delta &= (\bar{y}_{11} - \bar{y}_{12}) - (\bar{y}_{21} - \bar{y}_{22})
\end{aligned}$$

$$d = \frac{\mu_1 - \mu_2}{s}$$

$$s = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$$

$$n = \left(\frac{2 \cdot z \cdot \sigma}{W}\right)^2$$

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\delta}\right)^2$$

$$L_{h,i} = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2, & \text{if } |y_i - \hat{y}_i| \le \epsilon \\ \epsilon(|y_i - \hat{y}_i| - \frac{\epsilon}{2}), & \text{otherwise} \end{cases}$$

$$L_{ei,i} = \begin{cases} 0, & \text{if } |y_i - \hat{y}_i| \le \epsilon \\ |y_i - \hat{y}_i|, & \text{otherwise} \end{cases}$$

$$L_{sei,i} = \begin{cases} 0, & \text{if } |y_i - \hat{y}_i| \le \epsilon \\ (y_i - \hat{y}_i)^2, & \text{otherwise} \end{cases}$$

$$\eta = \text{const.}$$

$$\eta \to \frac{\eta}{k}$$

$$\eta = \frac{\text{const.}}{\left(1 + \frac{t}{k}\right)^{t_0}}$$

$$\eta = \frac{\text{const.}}{\beta\left(t + t_0\right)}$$

$$\begin{aligned}
\bar{y}_n &= \frac{1}{N_n} \sum_{i \in N_n} y_i \\
\text{MSE} &= \frac{1}{N_n} \sum_{i \in N_n} \left(\hat{y}_i - \bar{y}_n\right)^2 \\
J &= \frac{n_{left}}{N} \cdot \text{MSE}_{left} + \frac{n_{right}}{N} \cdot \text{MSE}_{right}
\end{aligned}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}_i\right)^2}$$

$$\begin{aligned}
\hat{y}_{diff} &= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right) \\
\text{Exp. Var} &= 1 - \frac{\sum_{i=1}^{n} \left(\left(y_i - \hat{y}_i\right) - \hat{y}_{diff}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}_i\right)^2}
\end{aligned}$$

$$\begin{aligned}
&F_m(X) = F_{m-1}(X) + \nu \hat{y}_m(X) \\
&\hat{y}_1(X) \\
&\hat{y}_2(X) \\
&\hat{y}_3(X)
\end{aligned}$$

$$w_{est.i} = \nu \log\left(\frac{1 - r_i}{r_i}\right)$$

$$\begin{aligned}
X &= USV^H \\
s^2 = &\text{ eigenvalues} \\
\text{columns of } U = &\text{ eigenvectors of } XX^H \\
\text{rows of } V^H = &\text{ eigenvectors of } X^H X \\
\text{with } &\frac{1}{N} \sum_{i=1}^{N} \hat{x}_i = 0
\end{aligned}$$

$$
\begin{aligned}
P_d &= X \cdot \left(V_d^H\right)^T \\
X_{rec} &= P_d \cdot V_d^H
\end{aligned}
$$

$$
\begin{aligned}
X &= W \cdot H \\
&X^{m\times n} \\
&H^{k\times n} \\
&W^{m\times k}
\end{aligned}
$$

$$
d_{Frob}(X, WH) = SE_{Frob.} = \frac{1}{2}\sum_{i,j}\left(X_{i,j} - (WH)_{i,j}\right)^2
$$

$$
\begin{aligned}
d_{Frob}(X, WH) \\
&+ \lambda\sum_{i,j}|W_{i,j}| \\
&+ \lambda\sum_{i,j}|H_{i,j}|
\end{aligned}
$$

$$
\begin{aligned}
d_{Frob}(X, WH) \\
&+ \frac{1}{2}\lambda\sum_{i,j}\left(W_{i,j}\right)^2 \\
&+ \frac{1}{2}\lambda\sum_{i,j}\left(H_{i,j}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&d_{Frob}(X, WH) \\
&+ r\left(\lambda\sum_{i,j}|W_{i,j}| + \lambda\sum_{i,j}|H_{i,j}|\right) \\
&+ (1-r)\left(\frac{1}{2}\lambda\sum_{i,j}\left(W_{i,j}\right)^2 + \frac{1}{2}\lambda\sum_{i,j}\left(H_{i,j}\right)^2\right)
\end{aligned}
$$

$$\delta_{i,j}^2 = (x_i - x_j)(x_i - x_j)^T$$
$$\hat{\delta}_{i,j}^2 = (\hat{x}_i - \hat{x}_j)(\hat{x}_i - \hat{x}_j)^T$$
$$Loss = \sum_{i<j} \left(\delta_{i,j} - \hat{\delta}_{i,j}\right)^2$$

$$x_i$$
$$\hat{x}_i$$
$$\delta_{i,j}^2$$
$$\hat{\delta}_{i,j}^2$$
$$\hat{X} = XV$$

$$Loss = \sum_{i=1}^{m} \left(\hat{X}_i - \sum_{j=1, i\neq j}^{m} W_{i,j} X_j\right)^2$$
$$\sum_j W_{i,j} = 1$$

$$Loss = \sum_{i=1}^{m} \left(Y_i - \sum_{i=1, i\neq j}^{m} W_{i,j} Y_j\right)^2$$
$$\text{with } W_{i,j} \text{ fixed}$$

$$Loss = \sum_{i=1}^{m} \sum_{l=1}^{s_i} \left(Y_i - \sum_{i=1, i\neq j}^{m} W_{i,j}^{(l)} Y_j\right)^2$$

$$p_{j|i;i\neq j} = \frac{\exp\left(-\frac{(\hat{x}_i - \hat{x}_j)^2}{2\sigma_i}\right)}{\sum_{k\neq i} \exp\left(-\frac{(\hat{x}_i - \hat{x}_k)^2}{2\sigma_i}\right)}$$
$$\sum_j p_{j|i} = 1 \forall i$$
$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2m}$$

$$PP(p) = 2^{-\sum_i p(i) \log_2 p(i)}$$
$$\sigma_i$$

$$q_{ij} = \frac{\frac{1}{1 + (\hat{y}_i - \hat{y}_j)^2}}{\sum_{k\neq i} \frac{1}{1 + (\hat{y}_i - \hat{y}_k)^2}}$$

$$KL(P,Q) = \sum_{i \neq j} p_{i,j} \log \left(\frac{p_{i,j}}{q_{i,j}} \right)$$

$$P \star = E_e$$

$$\begin{aligned} \text{Inertia} = \sum_{i=1}^{m} \min_{u_j \in C} \left[(x_i - \mu_j)^2 \right] \\ \mu_j \text{ cluster j centroid} \\ C \text{ set of clusters} \end{aligned}$$

$$\begin{aligned} &V = \frac{(1+\beta) \cdot \text{homogeneity} \cdot \text{completeness}}{\beta \cdot \text{homogeneity} + \text{completeness}} \\ &\beta \text{ - homog. } \leftrightarrow \text{ complete. trade-off} \end{aligned}$$

$$s = \frac{b-a}{max(a,b)}$$

$$L(\hat{\Theta}; \hat{X}, \hat{Z}) = p(\hat{X}, \hat{Z} | \hat{\Theta})$$

$$\text{argmax}_{\Theta} \left(E_{\hat{Z}|\hat{X},\hat{\Theta}} \log L(\hat{\Theta}; \hat{X}, \hat{Z}) \right)$$

$$\begin{aligned} \text{AIC} &= 2k - 2\ln(\hat{L}) \\ k &= \text{ num. param in model} \\ \hat{L} &= \text{ maximum likelihood function value} \end{aligned}$$

$$\begin{aligned} \text{BIC} &= k \ln m - 2\ln(\hat{L}) \\ k &= \text{ num. param in model} \\ \hat{L} &= \text{ maximum likelihood function value} \\ m &= \text{ sample size} \end{aligned}$$