NB: this is not a comprehensive list of formulae used in the course.

- Euclidean distance:

$$r_2(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2}$$

cf. $\mathrm{sim}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{1+r_2(\boldsymbol{x},\boldsymbol{y})}$ as a similarity measure

- Pearson correlation coefficient:

$$\rho(x, y) = \frac{1}{N-1} \sum_{n=1}^{N} \frac{(x_n - \mu_x)}{\sigma_x} \frac{(y_n - \mu_y)}{\sigma_y}$$

- Bayes Theorem

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

$$P(C_k|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_k)P(C_k)}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|C_k)P(C_k)}{\sum_{k=1}^{K} p(\boldsymbol{x}|C_k)P(C_k)}$$

- Bayes decision rule (cf. MAP decision rule)

$$k^* = \arg\max_k \; P(C_k \mid \mathbf{x}) \; = \; \arg\max_k \; P(\mathbf{x}|C_k)P(C_k)$$

- Naive Bayes for document classification
(vocabulary: $V = \{w_1, \ldots, w_{|V|}\}$, test document: $D = (o_1, \ldots, o_L)$)

  - Likelihood by Bernoulli document model

$$P(\boldsymbol{b}|C_k) = \prod_{t=1}^{|V|} [b_t P(w_t \mid C_k) + (1-b_t)(1-P(w_t \mid C_k))]$$

  - Likelihood by Multinomial document model

$$p(\boldsymbol{x}|C_k) \propto \prod_{t=1}^{|V|} P(w_t|C_k)^{x_t} \; = \; \prod_{i=1}^{L} P(o_i|C_k)$$

- Univariate Gaussian pdf:

$$p(x \,|\, \mu, \sigma^2) = N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

- Multivariate Gaussian pdf:

$$p(\mathbf{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right)$$

  Parameter estimation from samples:

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n \,, \qquad \hat{\boldsymbol{\Sigma}} = \underbrace{\frac{1}{N-1}}_{\text{NB: } N \text{ in case of MLE}} \sum_{n=1}^{N} (\mathbf{x}_n - \hat{\boldsymbol{\mu}})(\mathbf{x}_n - \hat{\boldsymbol{\mu}})^T$$

- Correlation coefficient:

$$\rho(x_i, x_j) = \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \,, \qquad \Sigma = (\sigma_{ij})$$

- Logistic sigmoid function:

$$y = g(a) = \frac{1}{1+\exp(-a)}$$
$$g'(a) = g(a)(1-g(a))$$

- Softmax activation function (for multiple output nodes):

$$y_k = \frac{\exp(a_k)}{\sum_{\ell=1}^{K} \exp(a_\ell)}$$