CSE 142 Midterm Information Sheet

Manhattan (L1) distance: $d(x,y) = \sum_{i=1}^{d} |x_i - y_i|$

Euclidian (L2) distance: $d(x,\ y) = \|x\ - y\| = \left(\sum_{i=1}^{d} (x_i - y_i)^2\right)^{1/2}$

Minkowski (Lp) distance: $d(x,y) = \left(\sum_{i=1}^{d} |x_i - y_i|^p\right)^{1/p}$

Laplace correction = $\dfrac{N_i+1}{|S|+k}$

m-estimate = $\dfrac{N_i+m\pi_i}{|S|+m}$

Sample mean: $\hat{\mu}_x = \frac{1}{n}\sum_i x_i$

Sample variance: $\hat{\sigma}_x{}^2 = \frac{1}{n}\sum_i (x_i - \hat{\mu}_x)^2$

Sample covariance: $\hat{\sigma}_{xy} = \frac{1}{n}\sum_i (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$

Sample covariance matrix: $\hat{\Sigma} = \frac{1}{k} X_z\, X_z{}^T \quad = \frac{1}{k} S$ where S is the scatter matrix

Minority class

$\quad \mathrm{Imp}(\dot{p}) = \min(\dot{p}, 1-\dot{p})$

Gini index

$\quad \mathrm{Imp}(\dot{p}) = 2\dot{p}(1-\dot{p})$

Entropy

$\quad \mathrm{Imp}(\dot{p}) = -\dot{p}\log_2(\dot{p}) - (1-\dot{p})\log_2(1-\dot{p})$

√Gini index

$\quad \mathrm{Imp}(\dot{p}) = \sqrt{2\dot{p}(1-\dot{p})}$

Proportion of positive instances in a (binary) data partition:

$\dot{p} = \dfrac{P}{P+N}$

In a k-class data partition:

$\dot{p}_i = \dfrac{C_i}{\sum_{i=1}^{k} C_i}$

Total impurity:

$\mathrm{Imp}(\{D_1, \dots, D_l\}) = \sum_{i=1}^{l} \frac{|D_i|}{|D|} Imp(D_i)$

Bayes Rule:

$$P(H_i \mid D) = \frac{P(D \mid H_i)\, P(H_i)}{P(D)}$$

False positive rate (FPR) $= \frac{FP}{N} = \alpha$

False negative rate (FNR) $= \frac{FN}{P} = \beta$

True positive rate (TPR) = $\frac{TP}{P}$ = Recall

True negative rate (TNR) = $\frac{TN}{N}$

Accuracy = $\frac{TP+TN}{P+N} = \left(\frac{P}{P+N}\right) TPR + \left(\frac{N}{P+N}\right) TNR$

Error rate = $\frac{FP+FN}{P+N}$

Precision = $\frac{TP}{\hat{P}}$

$$\text{F1 score} = \frac{2 \cdot precision \cdot recall}{precision + recall} = \frac{2 \cdot TP}{P + \hat{P}}$$

Scoring classifier margin: $z(x) = c(x)\, \hat{s}(x)$ (true class function * score)

Margin loss function: $L(z(x)) \rightarrow [0, \infty)$

Size of hypothesis space: $|\, H \,| = 2^{(\#\ \text{instances})}$

Ranking classifier error rate: *rank-err* = $^{err}/_{PN}$

Ranking classifier accuracy: *rank-acc* = 1 – *rank-err*

Min. training set size for PAC learning: $m \geq \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$

PAC learning outputs, with probability at least $1-\delta$, a hypothesis h such that $err_D < \varepsilon$

Algorithm GrowTree(D, F) – grow a feature tree from training data.

```
Input   : data D; set of features F.
Output  : feature tree T with labelled leaves.
if Homogeneous(D) then return Label(D);
S ←BestSplit(D, F) ;                    // e.g., BestSplit-Class (Algorithm 5.2)
split D into subsets D_i according to the literals in S;
for each i do
    if D_i ≠ ∅ then T_i ←GrowTree(D_i, F) ;
    else T_i is a leaf labelled with Label(D);
end
return a tree whose root is labelled with S and whose children are T_i
```

Multivariate least-squares regression
(homogeneous representation)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w} + \boldsymbol{\epsilon}$$

$$\begin{aligned}\hat{\boldsymbol{w}} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \\ &= \boldsymbol{S}^{-1}\boldsymbol{X}^T\boldsymbol{y}\end{aligned}$$

Least-squares minimization with regularization:

$$\boldsymbol{w}^* = \operatorname{argmin}\, (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) + \lambda\, r(\boldsymbol{w})$$

Algorithm Perceptron(D, η) – train a perceptron for linear classification.

```
Input   : labelled training data D in homogeneous coordinates; learning rate η.
Output  : weight vector w defining classifier ŷ = sign(w · x).
w ←0 ;                  // Other initialisations of the weight vector are possible
converged←false;
while converged = false do
    converged←true;
    for i = 1 to |D| do
        if y_i w · x_i ≤ 0          // i.e., ŷ_i ≠ y_i
        then
            w←w + η y_i x_i ;
            converged←false;       // We changed w so haven't converged yet
        end
    end
end
```