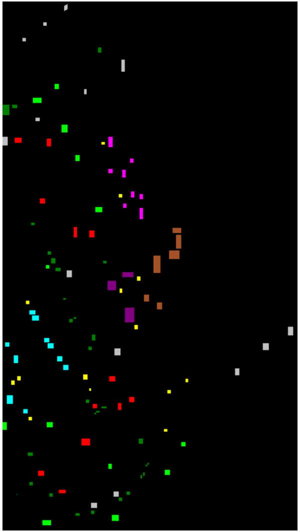
\>>> **Signal and Image processing in Remote Sensing**
\>>> **Classification**

Name: Mathieu Fauvel (UMR 1201 DYNAFOR INRA - INPT/ENSAT)
Date: 2016

>>> Classification of remote sensing images

Original Data



Ground-truth

Thematic Map

- Classification MAP: assign the pixel to the class with the highest probability.

$$\hat{y} = \max_{i=1,\ldots,C} p(y_i|\mathbf{x})$$

- Bayes rule: $p(y_i|\mathbf{x}) = p(\mathbf{x}|y_i)p(y_i)/p(\mathbf{x})$
- The decision rule becomes:

$$\max_{y=1,\ldots,C} p(\mathbf{x}|y_i)p(y_i)$$

- Gaussian model is conventionally used:

$$p(\mathbf{x}|y_i) = (2\pi)^{-d/2}|\Sigma_i|^{-1/2}\exp(-0.5(\mathbf{x}-\mu_i)^t\Sigma_i^{-1}(\mathbf{x}-\mu_i))$$

- $p(y_i)$ is usually approximated as the uniform distribution, i.e., $p(y_i) = 1/C$ or as the proportion of each class in the training set $p(y_i) = n_i/n$.
- Using the logarithm function and multiply by -2 the decision rule is :

$$k_i(\mathbf{x}) = (\mathbf{x}-\mu_i)^t\Sigma_i^{-1}(\mathbf{x}-\mu_i) + \ln(|\Sigma_i|) - 2\ln(p(y_i))$$

- ⋆ Maximization of the log-likelihood:

$$l = -2\ln(L) \propto n \ln\det(\boldsymbol{\Sigma}) + \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$

- ⋆ Derivate w.r.t $\boldsymbol{\mu}$:

$$\frac{\partial l}{\partial \boldsymbol{\mu}} \propto \sum_{i=1}^{n} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \Rightarrow \boxed{\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i}$$

- ⋆ Derivate w.r.t $\boldsymbol{\Sigma}$:

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} \propto n\boldsymbol{\Sigma}^{-1} - \sum_{i=1}^{n} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}$$

$$\Rightarrow \boxed{\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{\substack{i=1 \\ y_i = c}}^{n_c} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^t}$$

Detail of the matrix derivatives can be found in the matrix cookbook
http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3274

- Number of parameters to estimate for a d multidimensional Gaussian distribution. For the mean: d parameters. For the covariance matrix $d(d+1)/2$. So $d(d+3)/2$.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1d} \\ \vdots & \sigma_{22} & \dots & \sigma_{2d} \\ & & \ddots & \\ & & & \sigma_{dd} \end{bmatrix}$$

- Orthogonal matrix: $\mathbf{QQ}^t = \mathbf{I}$, hence

$$(\boldsymbol{\Sigma})^{-1} = \left(\mathbf{Q\Lambda Q}^t\right)^{-1} = \mathbf{Q\Lambda}^{-1}\mathbf{Q}^t = \sum_{i=1}^{d} \mathbf{q}_i\mathbf{q}_i^t \frac{1}{\lambda_i}$$

- Problem: find $\mathbf{A}$ such as $\hat{\mathbf{\Sigma}}\mathbf{A} = \mathbf{I}$ with $\hat{\mathbf{\Sigma}} + \varepsilon \Rightarrow \mathbf{A} + \varepsilon'$ (smothness condition)
- Minimization problem with penalization of non-smooth solutions:

$$\hat{\mathbf{A}} = \min_{\mathbf{A}} f(\mathbf{A}) \text{ with } f(\mathbf{A}) = \left\|\hat{\mathbf{\Sigma}}\mathbf{A} - \mathbf{I}\right\|^2 + \|\mathbf{\Gamma}\mathbf{A}\|^2$$

- Computing the derivative of f w.r.t. $\mathbf{A}$:

$$\frac{\partial f}{\partial \mathbf{A}} \propto \hat{\mathbf{\Sigma}}^t\hat{\mathbf{\Sigma}}\mathbf{A} - \hat{\mathbf{\Sigma}} + \mathbf{\Gamma}^t\mathbf{\Gamma}\mathbf{A}$$

- At the optimal, the derivative vanishes:

$$\hat{\mathbf{A}} = \left(\hat{\mathbf{\Sigma}}^t\hat{\mathbf{\Sigma}} + \mathbf{\Gamma}^t\mathbf{\Gamma}\right)^{-1}\hat{\mathbf{\Sigma}}$$

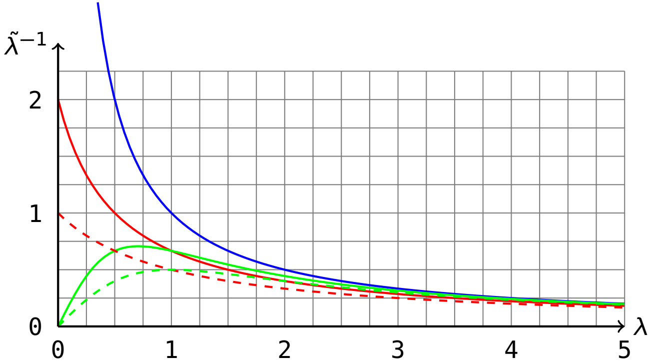
- Tikhonov : $\mathbf{\Gamma} = \alpha\mathbf{I}$

$$\hat{\mathbf{A}} = \left(\hat{\mathbf{\Sigma}}^2 + \alpha^2\mathbf{I}\right)^{-1}\hat{\mathbf{\Sigma}}$$

- Ridge : $\mathbf{\Gamma} = \alpha\hat{\mathbf{\Sigma}}^{1/2}$

$$\hat{\mathbf{A}} = \left(\hat{\mathbf{\Sigma}} + \alpha^2\mathbf{I}\right)^{-1}$$

- ★ Tikhonov: $\tilde{\lambda}_i^{-1} = \dfrac{\lambda_i}{\lambda_i^2 + \alpha^2}$
- ★ Ridge: $\tilde{\lambda}_i^{-1} = \dfrac{1}{\lambda_i + \alpha^2}$



>>> Support Vectors Machines

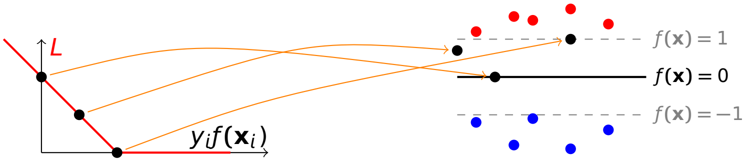
- ★ Supervised method: $\mathcal{S} = \left\{(\mathbf{x}_i, y_i)\right\}_{i=1}^{n}$, $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$

$$h(\mathbf{z}) = \operatorname{sign}\big(f(\mathbf{z})\big) \text{ with } f(\mathbf{z}) = \sum_{i=1}^{n} \alpha_i k(\mathbf{z}, \mathbf{x}_i) + b$$

- ★ Hyperparameters $\left(\{\alpha_i\}_{i=1}^{n}, b\right)$ learn by solving:

$$\min_{\boldsymbol{\alpha}, b} \left[\frac{1}{C} \|f\|^2 + \sum_{i=1}^{n} L\big(y_i, f(\mathbf{x}_i)\big) \right]$$

  - ★ $\|f\|^2 = \sum_{i,j=1}^{n} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$
  - ★ $L\big(y_i, f(\mathbf{x}_i)\big) = \max\big(0, 1 - y_i f(\mathbf{x}_i)\big)$



$$
\begin{array}{rcl}
\max_{\alpha} \; g(\alpha) & = & \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\
\texttt{constraint to} & & 0 \leq \alpha_i \leq C \\
 & & \sum_{i=1}^{\ell} \alpha_i y_i = 0
\end{array}
$$

Or

$$
\begin{array}{rcl}
\max_{\alpha} \; g(\alpha) & = & \boldsymbol{\alpha}^t \mathbf{1} - \frac{1}{2} \boldsymbol{\alpha}^t \left(\mathbf{K} \circ (\mathbf{y}\mathbf{y}^t) \right) \boldsymbol{\alpha} \\
\texttt{constraint to} & & 0 \leq \boldsymbol{\alpha} \leq \mathbf{1} \circ C \\
 & & \boldsymbol{\alpha}^t \mathbf{y} = 0
\end{array}
$$