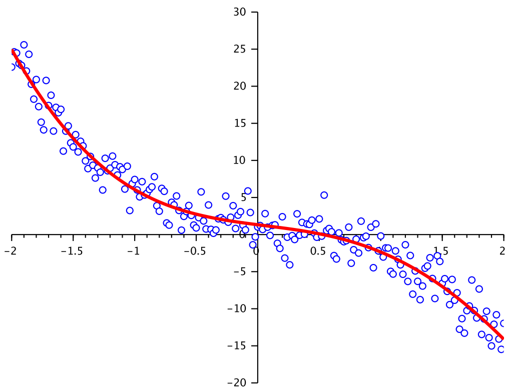
MT94/P24/TD - Problèmes de moindres carrés linéaires

1 Régression polynomiale avec validation



Le premier exercice de ce TD a pour but de vous initier à la méthode des moindres carrés ainsi qu'au principe de la validation.

On considère n couples $(t_i, y_i)_{i=1\ldots n}$ (les vecteurs `t` et `y` sont dans le fichier `data1.sod` disponible sur Moodle). Pour charger ce fichier dans Scilab et représenter les couples vous pouvez procéder ainsi :

```
--> load data1.sod;
--> plot(t,y,'o');
```

1.1 Régression

On cherche un polynôme de degré p défini par

$$P(t) = \theta_1 + \theta_2 t + \cdots + \theta_{p+1} t^p$$

tel que la quantité

$$S(\theta) = \sum_{i=1}^{n} (P(t_i) - y_i)^2,$$

est minimale, où on a noté $\theta = (\theta_1, \ldots, \theta_{p+1})^\top$.

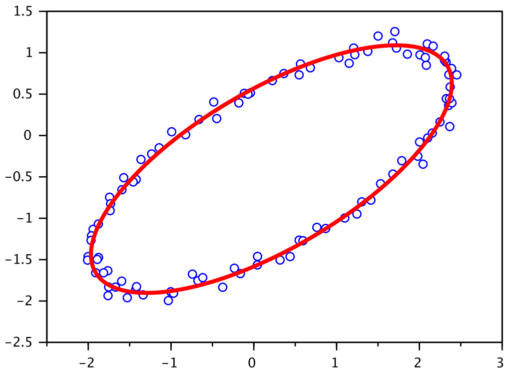
1. Montrez que
$$S(\theta) = \|A\theta - y\|^2,$$
où vous préciserez la matrice A.
2. A quelle condition sur la matrice A le problème de moindres carrés admet-il une solution unique ? A quelle condition sur les $(t_i)_{i=1\ldots n}$ cette condition est-elle vérifiée (on pourra faire des calculs pour $p = 1$ et $p = 2$ puis conjecturer. . .).
3. Ecrire un programme Scilab permettant de résoudre le problème de moindres carrés. Le programme doit représenter à l'écran les points $(t_i, y_i)_{i=1\ldots n}$ et le polynôme $P(t)$ obtenu pour des valeurs croissantes de son degré.

1.2 Validation

1. Tracer l'erreur de régression $S(\hat{\theta})$ en fonction du degré du polynôme. Que constatez-vous ? Cela permet-il de choisir le degré le plus adapté ?
2. Lisez la section "Model selection" du cours sur les moindres carrés et proposez une répartition des points $(t_i, y_i)_{i=1\ldots n}$ en un ensemble d'apprentissage $\mathscr{T}$ et un ensemble de validation $\mathscr{V}$.

3. On définit l'erreur d'apprentissage par
$$S_{\mathscr{T}}(\theta) = \sum_{i \in \mathscr{T}} (P(t_i) - y_i)^2,$$
et $\hat{\theta}^{(p)}$ les coefficients du polynôme de degré p rendant cette erreur minimale. Calculer et tracer $S_{\mathscr{T}}(\hat{\theta}^{(p)})$ en fonction de p (valeurs de p entre 1 et 8).
4. On définit l'erreur de validation par
$$S_{\mathscr{V}}(\theta) = \sum_{i \in \mathscr{V}} (P(t_i) - y_i)^2.$$
Superposer au tracé précédent $S_{\mathscr{V}}(\hat{\theta}^{(p)})$ en fonction de p et conclure.

2 Approximation d'une ellipse



Remarque : dans cet exercice les vecteurs et les matrices sont notés en caractères gras.

Une ellipse est composée des points du plan dont le vecteur de coordonnées $\mathbf{x} = (x, y)^\top$ vérifie l'équation implicite
$$(\mathbf{x} - \mathbf{c})^\top \mathbf{M} (\mathbf{x} - \mathbf{c}) = \gamma^2,$$
où $\mathbf{c} = (c_1, c_2)^\top$ est le centre de l'ellipse et
$$\mathbf{M} = \begin{pmatrix} 1 & \alpha \\ \alpha & \beta \end{pmatrix},$$
avec $\beta > \alpha^2$ (on pourra vérifier que $\mathbf{M}$ est définie positive). On dispose de points expérimentaux $\mathbf{x}_i = (x_i, y_i)$ où $i = 1 \dots n$, que l'on veut approcher au mieux par une ellipse en minimisant la distance algébrique
$$d(\alpha, \beta, \gamma, c_1, c_2) = \sum_{i=1}^{n} \left((\mathbf{x}_i - \mathbf{c})^\top \mathbf{M} (\mathbf{x}_i - \mathbf{c}) - \gamma^2 \right)^2.$$

1. En vous inspirant du cas du cercle (vu en cours), montrer que $d(\alpha, \beta, \gamma, c_1, c_2) = \|\mathbf{Ap} - \mathbf{b}\|^2$ où $\mathbf{A}$ est la matrice $n \times 5$ dont la ligne i est définie par
$$\mathbf{A}_i = \left(2x_i y_i,\, y_i^2,\, -2x_i,\, -2y_i,\, 1\right),$$
$\mathbf{b}$ est le vecteur de $\mathbb{R}^n$ défini par $b_i = -x_i^2$ et $\mathbf{p}$ est le vecteur défini par
$$\mathbf{p}^\top = \left(\alpha,\, \beta,\, c_1 + \alpha c_2,\, \alpha c_1 + \beta c_2,\, \mathbf{c}^\top \mathbf{M} \mathbf{c} - \gamma^2\right).$$
2. En déduire une méthode permettant de déterminer le vecteur $\mathbf{p}$ minimisant la distance algébrique, puis d'obtenir dans l'ordre α, β (pour ces deux là c'est facile), le vecteur $\mathbf{c}$, puis γ.
3. Appliquez cette méthode aux points $(x_i, y_i)_{i=1\dots n}$ (les vecteurs `x` et `y` sont dans le fichier `data2.sod` disponible sur Moodle).
4. Tracez l'ellipse obtenue (on pourra utiliser la macro `spec` de Scilab pour calculer les valeurs propres et vecteurs propres de $\mathbf{M}$, qui permettent d'obtenir facilement un paramétrage).