Université Paris I, Panthéon - Sorbonne

Master 2 E.T.E. 2013 – 2014

TP de Séries temporelles n^0 4: Prédiction pour une série chronologique

Dans ce TP, on utilise des techniques de régression (globale ou locale) pour prédire les valeurs futures d'une série chronologique. On compare avec les prédictions optimales par moindres carrés (calculées théoriquement). On utilise également un filtrage exponentiel.

Prédiction par régression polynômiale

x=1:20	On génère la variable temps.
y=(-1)^x+x–(0.1*x^2) + 5*rnorm(20)	On génère une série chronologique y (sous la forme d'un vecteur). Décrire théoriquement la série (tendance, saisonnalité et bruit).
y.reg=glm(y ∼ poly(x,3))	Régression polynômiale de degré 3 de y par x. La commande *glm* généralise la commande lm précédemment utilisée. Elle permet notamment des régressions pondérées (voir plus bas).
plot(x,y)	Visualisation de la série.
lines(x,y.reg$fit)	Graphe du polynôme obtenu par la régression.
new=data.frame(x=21)	On génère un tableau de données contenant la valeur ou les valeurs de x pour lesquelles on va faire une prédiction.
y.pred=predict(y.reg,new,se=TRUE)	Prédiction au temps $x = 21$ obtenue par le polynôme.
y.pred	La commande $se = TRUE$ permet d'avoir une estimation de l'écart--type de cette prédiction (par défaut elle n'est pas donnée; essayez...). En remplaçant 21 par $c(25, 30, 50)$, on obtient les prédictions aux temps $x = 25,\ 30$ et 50. Comparer (donner des écarts numériques) avec la prédiction optimale par moindres carrés (ce que l'on aurait obtenu si le modèle théorique exact était identifié). Refaire la même chose avec un polynôme de degré 1, 2 et 10. Quel est votre choix (théorique et empirique) quant au degré de la régression?

Prédiction par lissage exponentiel

Cette fois-ci on pondère la régression polynômiale précédente par des poids de type exponentiel.

wei=1^x	Génération d'un vecteur poids (ici uniforme).
y.regp1=glm(y ∼ poly(x,2),weights=wei)	Régression polynômiale de degré 2 de y par x, pondérée par *wei*. Quelle est la différence avec la régression précédente de degré 2? Expliquer ce résultat. Générer un poids ne portant que sur la première donnée, puis un poids ne portant que sur la dernière. Prédire pour $x = 25$. Expliquer ces résultats.
wei=0.8^(21-x)	Génération d'un vecteur poids exponentiel. Prédire pour $x = 25$. Faire varier le paramètre $\beta = 0.8$. Quelle valeur est-elle préférable? (c'est-à-dire, quelle est la valeur (à 0.1 près) permettant d'obtenir la meilleure prédiction?)

Prédiction à partir d'une régression locale

La commande *loess*, qui permet des estimations par régression locale, permet également d'obtenir des prédictions. En revanche, cela n'est pas le cas pour la commande *lowess*, pourtant plus robuste.

`y.lo1=loess(y~x,co=loess.control(surface="direct"))`	Régression *loess*, à laquelle on adjoint une option qui permettra de faire des prédictions sans limitation.
`plot(x,y)`	Tracé des données.
`lines(x,y.lo1$y)`	Tracé du résultat de la régression locale.
`y.lo2=loess(y~x,span=0.4,co=loess.control(surface="direct"))`	Régression *loess* dans laquelle on explicite la taille de la fenêtre locale, qui par défaut vaut 0.75.
`lines(x,y.lo2$y)`	Tracé du résultat de la seconde régression locale. Comparer les deux tracés. Faire d'autres essais avec d'autres valeurs de *span*.
`new=data.frame(x=seq(19,25,0.5))`	On génère une suite de nombres que l'on transforme en un tableau de données. Comment fonctionne la commande *seq* (essais...) ?
`y.prelo1=predict(y.lo1,new,se=TRUE)`	On utilise la première régression *loess* pour obtenir une prédiction pour les valeurs de x données dans le tableau *new*. Tracer sur une même fenêtre les différentes prédictions que vous obtenez en changeant *span*. Sur le même dessin, rajouter la prédiction optimale par m.c. (théorique) en ces points. Enfin, comparer avec les deux méthodes vues précédemment.

Exercice

Reprendre la série *hstart*. Prédire le nombre de déménagements pour les douze mois de 1975 (n'oubliez pas la saisonnalité!). Détailler et expliquer la démarche choisie. Refaire la même chose en travaillant à partir des données de 1960 à 1973 (estimation de la tendance et de la saisonnalité), puis prédire à l'aide d'un filtrage exponentiel les 12 mois de 1974. Quel est l'intérêt d'un tel travail?