

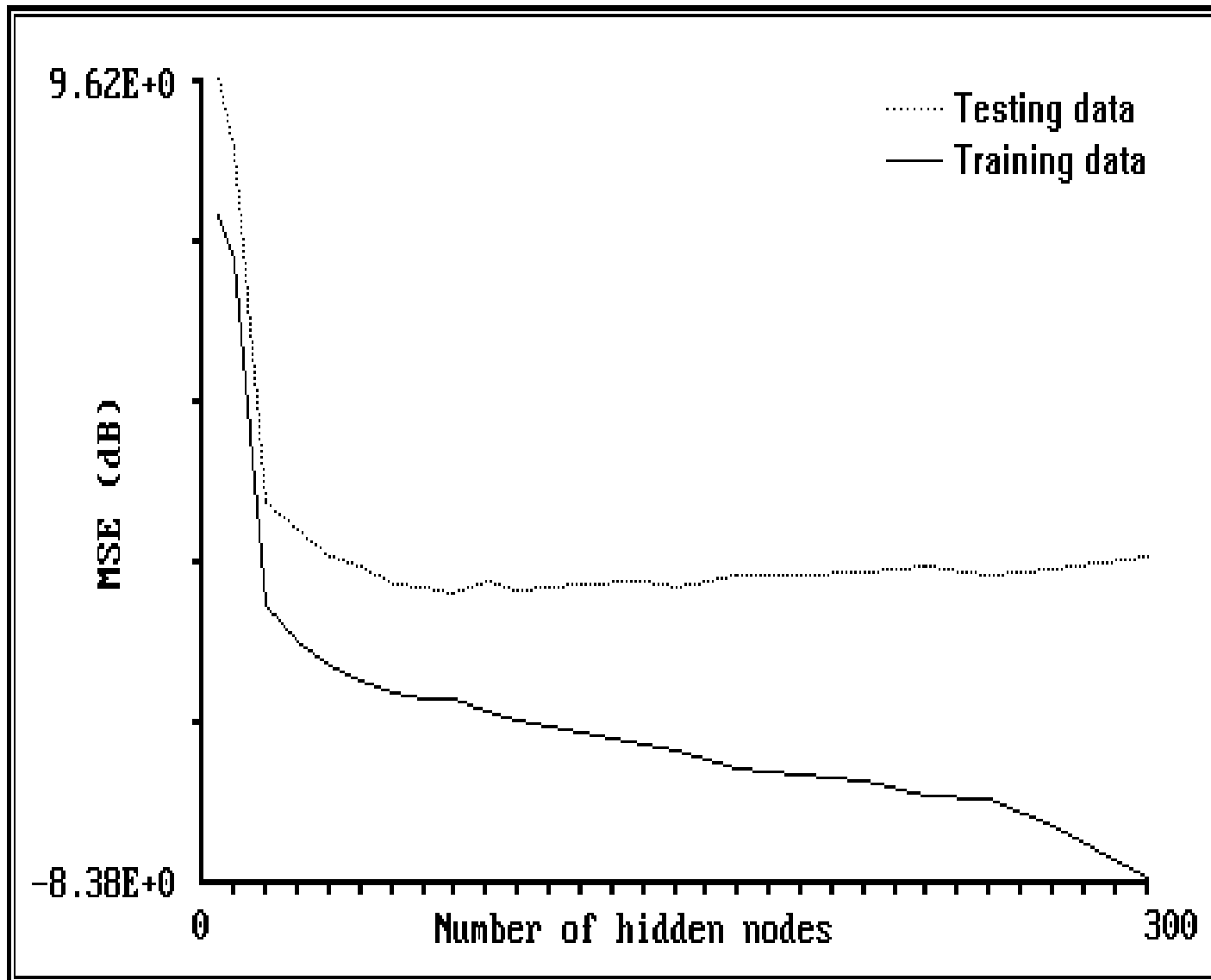
Erreur d'apprentissage
Erreur de généralisation
Validation croisée

Qu'est-ce que l'erreur d'apprentissage?

- On entraîne l'algo d'apprentissage avec notre ensemble d'entraînement et ensuite nous testons l'algo avec **le même ensemble de données**
- Qui a-t-il de très mal à faire ça?

Qu'est-ce que l'erreur d'apprentissage?

- On entraîne l'algo d'apprentissage avec notre ensemble d'entraînement et ensuite nous testons l'algo avec **le même ensemble de données**
- Qui a-t-il de très mal à faire ça?
 - L'erreur que nous calculons est biaisée, car l'algo est devenu spécialiste de l'ensemble d'entraînement
 - L'erreur calculée ne dit absolument rien des capacités qu'aura l'algo sur de nouvelles données



Nous sommes plutôt intéressés par l'erreur de généralisation

- L'erreur de généralisation = l'erreur sur de nouvelles données
- Habituellement nous avons 2 ensembles de données un d'entraînement et un de test
 - Il est donc possible de calculer l'erreur de généralisation avec l'ensemble de test

Nous sommes plutôt intéressés par l'erreur de généralisation

- L'erreur de généralisation = l'erreur sur de nouvelles données
- Habituellement nous avons 2 ensembles de données un d'entraînement et un de test
 - Il est donc possible de calculer l'erreur de généralisation avec l'ensemble de test
- Que faire quand nous n'avons pas beaucoup de données et que l'algo a besoin de « toutes » les données?

Nous sommes plutôt intéressés par l'erreur de généralisation

- L'erreur de généralisation = l'erreur sur de nouvelles données
- Habituellement nous avons 2 ensembles de données un d'entraînement et un de test
 - Il est donc possible de calculer l'erreur de généralisation avec l'ensemble de test
- Que faire quand nous n'avons pas beaucoup de données et que l'algo a besoin de « toutes » les données? Utiliser la validation-croisée (cross-validation)

Principe général de la validation croisée

- Diviser toutes les données disponibles en k groupes
- Erreur = 0;
- Pour i allant de 1 à k
 - ensemble_Test = groupe(i) ;
 - ensemble_Entraînement = groupe(1 à $k - i$);
 - AA(ensemble_Entraînement);
 - Erreur += erreur_test_AA(AA,ensemble_Entraînement);

Validation croisée « Leave-one-out »

- Supposons que nous disposons de n données, alors $k = n$ et nous avons des groupes de dimension 1
- Erreur = 0;
- Pour i allant de 1 à n
 - ensemble_Test = $N(i)$;
 - ensemble_Entraînement = $N(1 \text{ à } k - i)$;
 - AA(ensemble_Entraînement);
 - Erreur += erreur_test_AA(AA,ensemble_Entraînement);