Université de Jijel
Département d'Electronique
Master 1 : Microélectronique

Travaux pratiques N°03

Modélisation de la tension de seuil V_{th} par régression linéaire multiple

En microélectronique, la tension de seuil V_{th} d'un transistor MOS est sensible aux variations du procédé de fabrication. On souhaite établir un modèle prédictif simplifié pour estimer V_{th} sans avoir à résoudre les équations physiques complexes à chaque fois.

On considère trois paramètres d'entrée :

1. T_{ox} (nm) : Épaisseur de l'oxyde de grille.
2. N_A (10^{17} cm^{-3} : Concentration de dopage du substrat.
3. V_{sb} (V) : Tension de polarisation du substrat.

Le tableau suivant présente les mesures effectuées sur 8 échantillons de test :

Échantillon	T_{ox} (nm)	NA (10^{17} cm^{-3})	V_{sb} (V)	V_{th} mesuré (V)
1	2.1	1.5	0.0	**0.40**
2	2.5	2.0	0.5	**0.55**
3	3.0	1.8	1.0	**0.65**
4	2.2	2.5	0.2	**0.52**
5	2.8	3.0	0.8	**0.72**
6	3.5	0	0.1	**0.50**
7	100	4.0		**0.85**
8	3.2	2.2	0.6	**0.68**

Travail demandé

1. Écrire l'équation théorique du modèle de régression linéaire multiple reliant V_{th} aux trois paramètres de fabrication. Définir chaque terme ($\theta 0$, $\theta 1$, $\theta 2$, $\theta 3$).

2. Séparer et identifier les variables indépendantes, la variable dépendante, les valeurs : aberrante, nulle (NA), manquante (NaN).

3. En utilisant Python et le **framework** *Scikit-Learn*, développez un script permettant de :

   - Charger les données dans un **DataFrame**, puis les affiche.
   - Nettoyer les données de la table (utiliser : la suppression, la médiane, la mode ou la moyenne).
   - Afficher et explorer les données de la table.
   - Choisir le modèle de régression : sélectionner les variables indépendantes (feature) et la variable dépendante (target). Puis, initialiser le modèle choisi.
   - Entrainer le modèle : séparer les données en ensemble d'entrainement et de test (80%, 20%), avec random_state=42.). Puis, entrainer le modèle avec les données d'entraînement.

- Afficher l'intercept et les coefficients du modèle.
- Afficher l'équation finale du modèle.
- Comparer les valeurs réelles et prédites.
- D'après les coefficients obtenus, quel paramètre semble avoir l'impact le plus important sur la tension de seuil ?
- Calculer le coefficient de détermination (R^2) et l'erreur quadratique moyenne (RMSE). Ensuite, interpréter la qualité du modèle obtenu.
- Un nouveau lot de fabrication présente les caractéristiques suivantes : T_{ox} = 2.4 nm, NA = 2.1 10^{17} cm^{-3} et V_{sb} = 0.3V. Quelle est la valeur de V_{th} prédite ?

- Changer le pourcentage de séparation des données. Que se passe-t-il si on change le pourcentage de séparation entre l'ensemble d'entraînement et l'ensemble de test ? Par exemple, (70%,30%) et (90%,10%).
- Changer la valeur de random_state : Comment le modèle change-t-il si on essaie différentes valeurs pour random_state ? Par exemple, utiliser 12, 22 ou d'autres valeurs.
- Introduire la validation croisée (cross-validation) : Quelle est la performance du modèle avec une validation croisée (par exemple, avec 5 ou 10 plis (folds)) ? Comment cela affecte-t-il la stabilité des résultats?
- Ajouter des interactions entre les variables : Comment les résultats changent-ils si on introduit des termes d'interaction entre les variables ? Par exemple, ajouter une interaction entre (Tox) et (NA) dans le modèle : $T_{ox}\_N_A = T_{ox} \times N_A$.
- Analyser l'impact des valeurs aberrantes : Quel est l'impact des valeurs aberrantes sur le modèle ? Par exemple, que se passe-t-il si on laisse l'échantillon avec (Tox = 100) dans le jeu de données ?
- Même question pour les valeurs manquantes et les valeurs nulles.
- Analyser la corrélation entre les variables indépendantes : Quelle est la corrélation entre les différentes variables indépendantes (Tox), (NA), et (Vsb) ? Est-ce qu'il y a des problèmes de multicolinéarité ?