Data drift refers to the changing distribution of the data to which the model is applied. Concept drift refers to a changing underlying goal or objective for the model.

Data drift, or covariate shift, refers to the phenomenon where the distribution of data inputs that an [ML model](https://www.aporia.com/learn/machine-learning-model/machine-learning-challenges-and-solutions/) was trained on differs from the distribution of the data inputs that the model is applied to.

Concept drift occurs when there is a change in the functional relationship between a model’s input and output data. The model continues to function the same despite the changed context, unaware of the changes. Thus, the patterns it has learned during training are no longer accurate.

 **Concept Drift**: Se le preferenze dei consumatori cambiano radicalmente a causa di una nuova moda o tendenza, il modello potrebbe non essere più accurato nel prevedere le vendite. In questo caso, la relazione tra le caratteristiche di input (demografia, marketing) e l'output (vendite) è cambiata.

 **Data Drift**: Se la composizione demografica dei clienti cambia nel tempo (ad esempio, una maggiore percentuale di giovani rispetto agli anziani), il modello potrebbe non performare bene se addestrato su dati vecchi. Qui, la distribuzione delle caratteristiche di input è cambiata, anche se la relazione tra input e output potrebbe rimanere la stessa.

X training data

F learned model on X

X’ inference data

X’ differisce da X quindi F(X) != F(X’), modello trainato su dati con distribuzione molto diversa

DATA DRIFT

Per detectare il data drift(cambia la distribuzione dei dati su cui usando il modello allenato devo fare inferenza, i nuovi dati su cui faccio inference però sono di una distribuzione diversa rispetto a quelli su cui ho allenato il modello quindi c’è data drift:

DATA DRIFT DETECTION

* **Monitoraggio della Distribuzione dei Dati**: Utilizzare test statistici come il test di Kolmogorov-Smirnov per rilevare cambiamenti nella distribuzione dei dati. Is there a statistically significant difference between the distribution of the recent X and the one of the old X on which I have trained the model?
* **Feature Importance Tracking**: Monitorare l'importanza delle caratteristiche per verificare se alcune diventano meno rilevanti nel tempo.

**Per detectare invece il concept drift**: cambia la relazione fra input(x) e output, la funzione stessa cambia:

**CONCEPT DRIFT DETECTION**

* **Monitoraggio delle Prestazioni del Modello**: Misurare costantemente l'accuratezza e altre metriche di performance del modello.
* **Adaptive Learning**: Utilizzare modelli che possono adattarsi dinamicamente ai nuovi dati.(ADWIN)

**X = {ir\_0,ir\_1,ir\_2,ir\_3** …, ir\_6}

Y = 0,1,2 (vai a destra, sinistra o dritto)

**Cosa potrei fare**:

batch ml

- traino con condizioni di bianco su nero: testo con circuito diverso ma sempre bianco su nero, il modello sarà in grado di muoversi✔️

- Cambio le condizioni quindi uso il modello ad esempio variando il colore, ad un certo punto non è in grando di andare ✔️

Streaming ML

Anche nel caso di un drift devo essere in grado di avere il ground truth

Data drift: cambia la distribuzione dei dati, sto testando su degli input dei sensori che hanno una distribuzione diversa da quella su cui ho trainato.

Concept drift: cambia la relazione se prima era di un certo tipo, ora