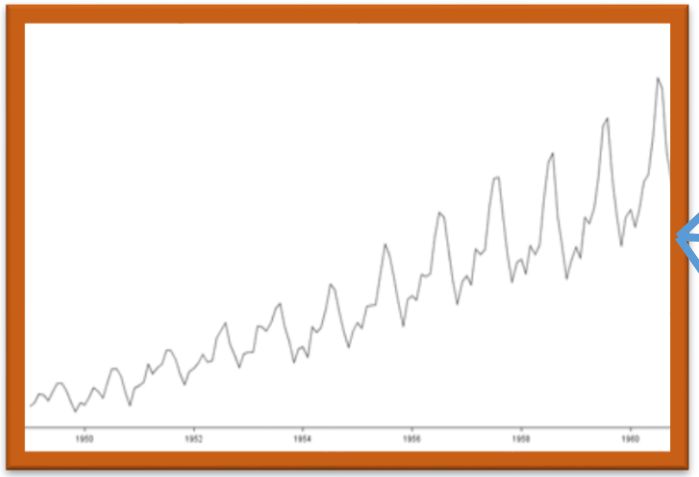
Model Quality Criteria

A (VERY) Brief Overview

Model Quality Criteria

Model Ensembles



Model 1

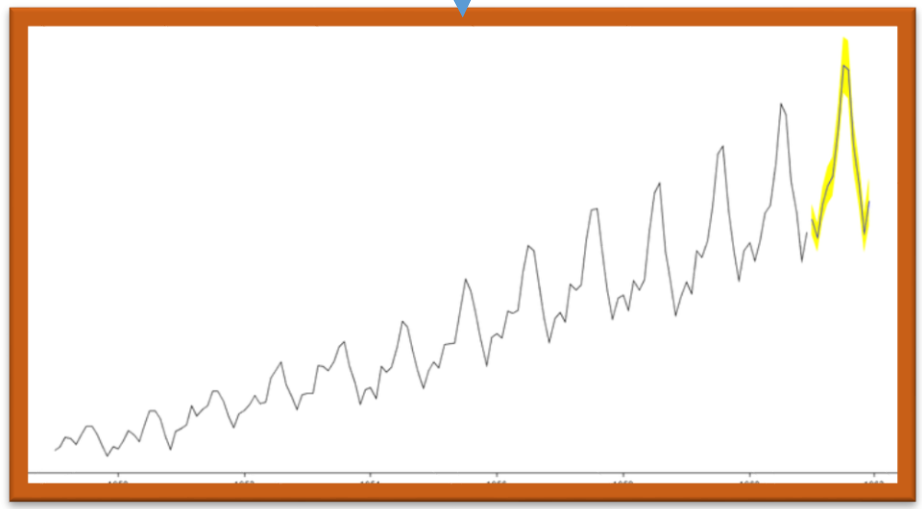
Model 2

Model N

w_1

w_2

w_N



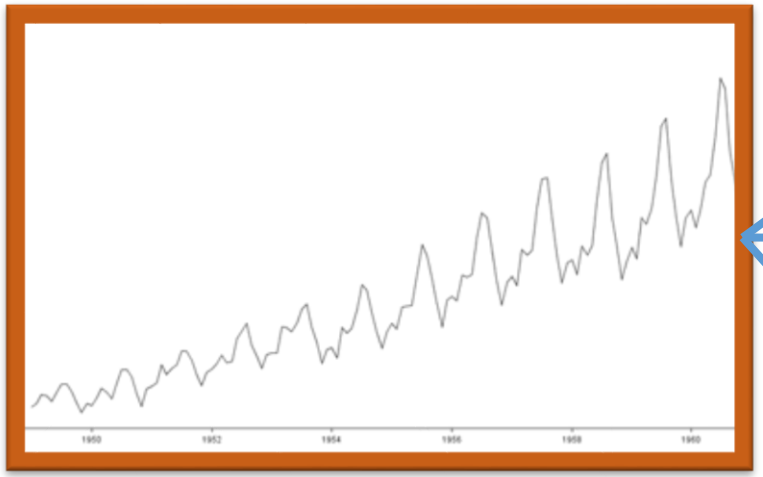
Compute the weighted sum of all models

??? How to find the w_{M_j} ???

- $w_{M_j} = \frac{1}{N} \Rightarrow M = \frac{\sum_{j=1}^{N} M_j}{N}$
- $w_{M_j} = argmin\left(Cost(\varepsilon)\right)$
- … … …

Model Quality Criteria

Model Competitions



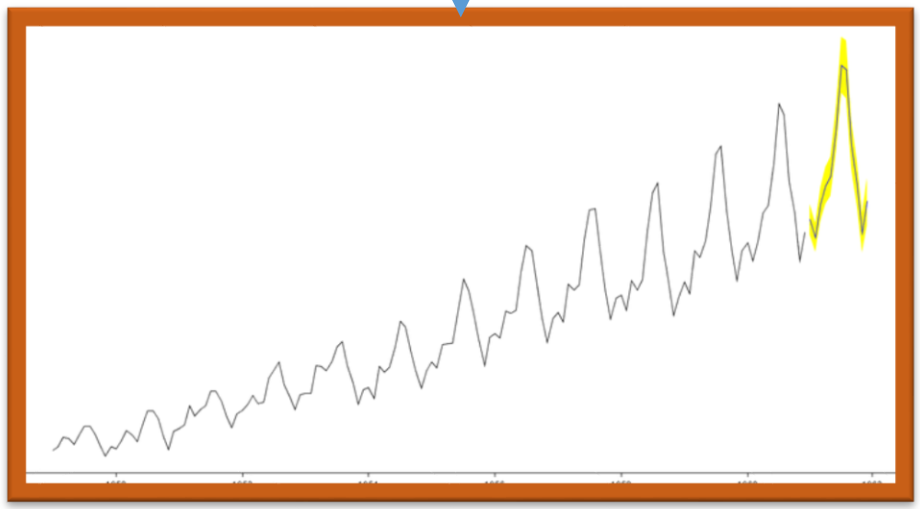
Only use the best model

Model 1

Model 2

Model N

$Quality_{M_1} < Quality_{M_2}$?



??? How to quantify the $Quality_{M_j}$???

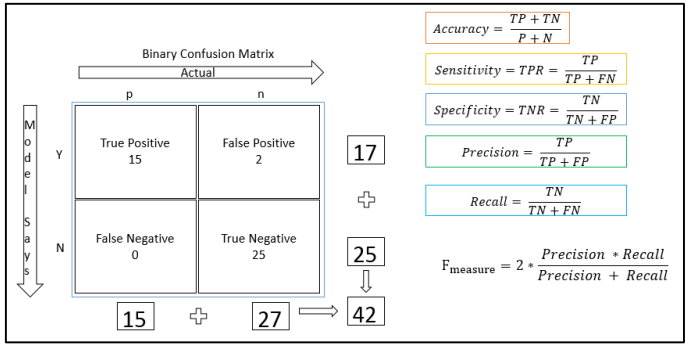
Model Quality Criteria

Specific Criteria: What question does the model answer?

Prediction:

Within the scope of ***X***?

Outside the scope of ***X***?



Latent Dependencies & Features?

χ^2 -based criteria:

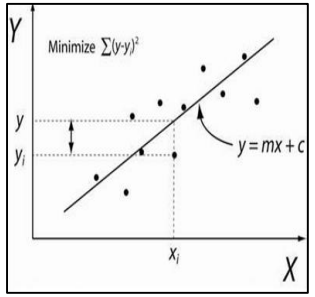
$$NFI = 1 - \frac{\chi_M^2}{\chi_B^2}, \quad where\ \mathrm{M} = model; \mathrm{B} = baseline$$

$$CFI = 1 - \frac{\delta_M}{\delta_B}, \quad where\ \delta = \max(\chi^2 - df, 0)$$

Etc.

Behavior Explanation?

$$R^2 \equiv 1 - \frac{Var(Y_{predicted} - Y_{observed})}{Var(Y_{observed})}$$

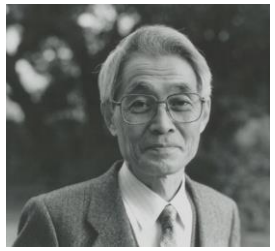


Sensitivity?

$$P\{a_i \mid X, Y\}$$

Model Quality Criteria

Universal (task-agnostic) Criteria



$$AIC = 2 * K - 2 * \ln(L)$$

$$AIC_c = 2 * K - 2 * \ln(L) + \frac{2 * K * (K+1)}{n - K - 1}$$

$$BIC = 2 * K * \ln(n) - 2 * \ln(L)$$

??? Which one to use ???

- Task-specific
- A list of criteria and values
- $Q_M = \sum_{i=1}^{N_c} w_i * Q_i$
  - $Q_i = scaled\ quality\ criterion\ \#i$
  - $w_i = weight\ (importance)\ of\ Q_i$

Mean squared error	$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$
Mean absolute percentage error	$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n} \left\lvert \frac{e_t}{y_t} \right\rvert$