k-NN Δ-Forecast for El Nino 3.4 Anomalies

dara@lossofgenerality.com

Data

1950	1	24.84	26.26	-1.42
1950	2	25.22	26.53	-1.31
1950	3	26.04	27.09	-1.04
...				
2013	10	26.65	26.79	-0.14
2013	11	26.54	26.74	-0.20
2013	12	26.20	26.69	-0.49

Δ

$\Delta_t = X_t - X_{t-1}$

Example: let t be 1950,2 then

$\Delta_{t=1950,2} = -1.31 - (-1.42) = 0.11$

2013, 12:

$\Delta_{t=2013,12} = -0.49 - (-0.20) = -0.29$

0.11
0.27
-0.08

...
0.14
-0.06
-0.29

Forecast Model

$X_{t+1} = X_t + \text{Forecast}(\Delta_t)$

$\Delta_{t+1} \approx \text{Forecast}(\Delta_t)$ since:

$\Delta_{t+1} = X_{t+1} - X_t \implies X_{t+1} = X_t + \Delta_{t+1}$

This model avoids wobbling and oscillation and reduces the over-fitting but prone to noise. For noisy data Wavelet Transforms which remove the noise create near ideal forecast for Trend.

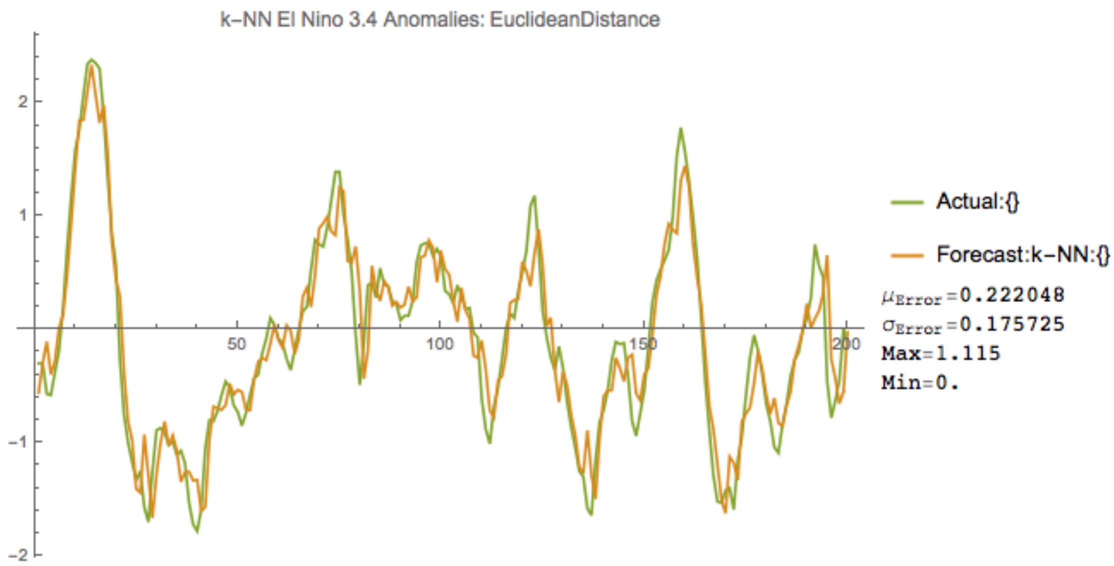
k-NN

http://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.

Next Month Forecast

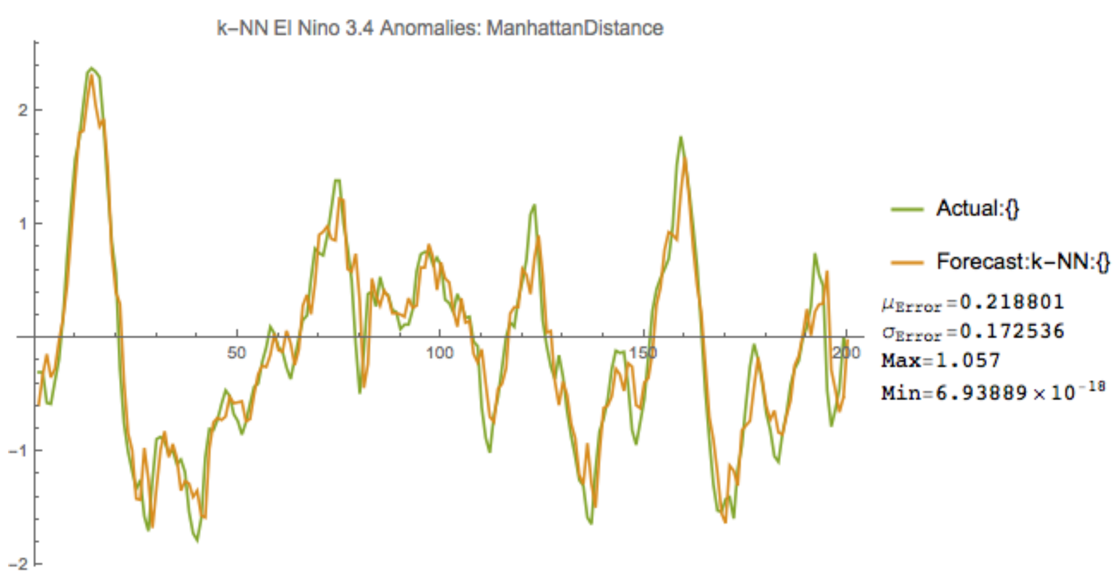
The El Nino 3.4 Δ data was truned into sequential list of vectors of length 6, and then 10 Nearest-neighbour vectors computed and then their corresponding next month value i.e. at next month was averaged.

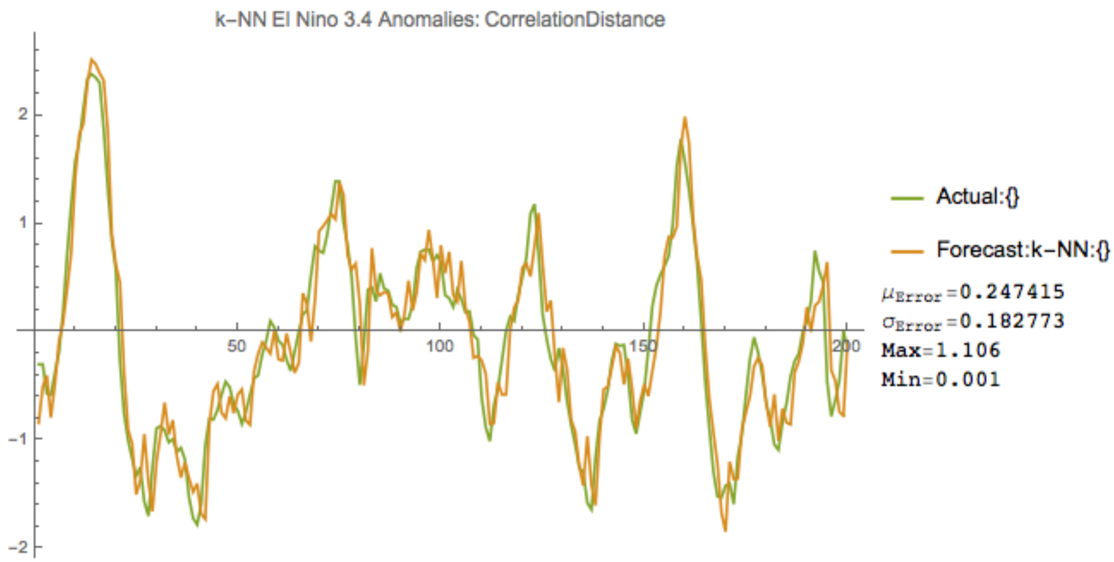


Best Metric turned out to be Manhattan Distance

$d((a, b, c), (x, y, z)) = \text{Abs}(a - x) + \text{Abs}(b - y) + \text{Abs}(c - z)$

Thefore the best forecast was issued by non-Euclidean metric with %9.1 error.





6th Month Forecast

The El Nino 3.4 Δ data was truned into sequential list of vectors of length 6, and then 10 Nearest-neighbour vectors computed and then their corresponding next month value i.e. at next 6th month was averaged.

Best forecast turned to be with Canberra Distance with %8.4 accuracy:

$$d((a, b, c), (x, y, z)) = \frac{\text{Abs}(a-x)}{\text{Abs}(a)+\text{Abs}(x)} + \frac{\text{Abs}(b-y)}{\text{Abs}(b)+\text{Abs}(y)} + \frac{\text{Abs}(c-z)}{\text{Abs}(c)+\text{Abs}(z)}$$