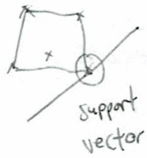
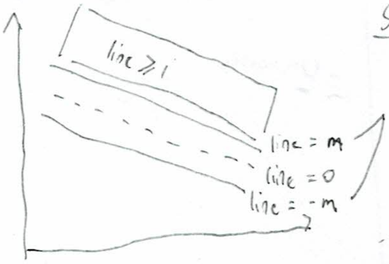
SVM Classification

line: $a_1x_1 + a_2x_2 + \dots + a_mx_m + a_0 = 0$

$$\sum_{i=1}^{m} a_ix_i + a_0 = 0$$





Scaling data recomended

Distance btw. 2 parallel lines/Margin:

$$= \frac{2}{\sqrt{\sum_i (a_i)^2}}$$ Maximize Margin

Minimize $\sum_i (a_i)^2$ Margin

Soft classifier Error:

$$\left(\sum_{i=1}^{m} a_ix_{ij} + a_0\right) y_j - 1 < 0$$

$$\text{minimize} \left| \max\left\{0,\ 1 - \left(\sum_{i=1}^{m} a_ix_{ij} + a_0\right) y_j\right\} \right| \text{error} + \lambda \sum_{i=1}^{m} (a_i)^2$$

weight

SVM sensitive to data scale differences as scale affects line coefficient, small changes to data point will affect line significantly.

$0 = 700 + \boxed{3}x_1 + \boxed{1500}x_2$

Scaling

Scaling: 0-1 $\dfrac{x_{ij} - x_{min}}{x_{max} - x_{min}}$

Standardization: To normal dist.

$\mu = 0$, s.d. $= 1$

$$\frac{x_{ij} - \mu}{\sigma}$$

K-Nearest Neighbour

dist $= \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2 (w_i)}$ straight line

weight for dimension

Manhatten dist $= |x_1 - y_1| + |x_2 - y_2|$

Generalize 2 Methods $= \sqrt[p]{|x_1 - y_1|^p + |x_2 - y_2|^p}$

$$= \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}$$

p-norm dist.

∞-norm dist. = sum of largest $|x_i - y_i|$ to infinity power

$$= \sqrt[\infty]{\max |x_i - y_i|^\infty}$$

$$= \max |x_i - y_i|$$

K-nearest Neighbour.

Heuristic: Fast but not guaranteed to be best

$$\text{Minimize}_{y, z} \sum_i \sum_k y_{ik} \sqrt{\sum_j (x_{ij} - z_{jk})^2}$$

y_{ik}: data point in cluster 1, 0

x_{ij}: data point

z_{jk}: cluster center

Change Detection - CUSUM.

$S_t = \max\{0,\ S_{t-1} + (x_t - \mu)\}$

$S_t = \max\{0,\ S_{t-1} + (x_t - \mu - C)\}$

C: weight.

$S_t \geq$ Threshold?

$S_t \geq$ Threshold

$(\mu - x_t - C)$ ...

Exponential Smoothing

Time Series Data.

$S_t = ax_t + (1-a)S_{t-1}$

S_t: Expected, x_t: observed, a: weight (0-1)

High Randomness = ↓ a we estimate S_{t-1} dont trust observed x

$S_t = ax_t / C_{t-L} + (1-a)(S_{t-1} + T_{t-1})$

Single/Double/Triple Exp. Smoothing.

Winter Holts. Method.

$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$

L: Length of cycle.

Forecast

$F_{t+1} = (S_t + T_t)C_{(t+1)-L}$

Error

$\min\{F_t - x_t\}$

Exponential

$S_t = ax_t + (1-a)ax_{t-1} + (1-a)^2 ax_{t-2} + (1-a)^3 ax_{t-3} + \dots$

Uses all data.

ARIMA

Forecasting with Exponential Smoothing if μ, s.d., measures are constant overtime.

Key Differences:

1. Order of differences. D^{th} order. $(x_t - x_{t-1})$.
2. Prediction based on previous values (Autoregressive)

autoregression on differences.

order-p autoregressive: how far back to consider.

Exponential Smoothing = ∞ order, all the way back.

3. Using error back to q^{th} order back.

GARCH

Generalized. Auto Regressive. Conditional Heteroskedasticity

Estimate / Forecast the Variance.

$e^{(AIC1 - AIC2)}$

≈ Model 2 is _ than 1

Transformation

1. Box. Cox: Reduce Heteroskedasticity (Homogenity of variance.)

$t(y) = (y^{\lambda} - 1)/\lambda$ Log-transform.

2. Detrending: Time series data. Fit simple regression and subtract away from actual value.

price = observed − $y(x)$

PCA.

- More spread data explains more variance.
- Scale value that $\mu = 0$
- linear transformation, XV_1, XV_2, V: Eigenvector.