in the parameters \( \theta \) is given. If the samples in \( C \) give no information

Assumption:

\[
p(x|w, \theta) = \frac{1}{Z} e^{\langle x, w \rangle \theta}
\]

\( \theta = \mathcal{N}(0, I_k) \) is parameterized by \( \Theta \)

\( p(x|w) \) is parametrized by \( \Theta \)

The samples are i.i.d. \( \mathcal{N}(0, I_k) \)

Samples in \( C \) drawn according to \( p(x|w) \)

\( c \) data sets \( 1, \ldots, n \) with

ML Estimators

- Maximum likelihood
- MAP
- SVM
- \( \mathcal{L} \) Minimization
- \( \mathcal{L} \) Minimization
- \( \mathcal{L} \) Minimization
- \( \mathcal{L} \) Minimization
- \( \mathcal{L} \) Minimization
- \( \mathcal{L} \) Minimization

Recall

Lecture 3
The observation vectors $\tilde{x}_1, \ldots, \tilde{x}_n$ maximize $
abla_{\theta} \log p(x|\theta)$

where $\arg \max$ is

In some sense, best guesses with respect to the data

That is, $\theta$ corresponds to the value of $\theta$ that maximizes $p(\tilde{x}|\theta)$

Vector $\theta$ that maximizes $p(\tilde{x}|\theta)$ is the ML estimate of $\theta$.

Definition: 

The samples $\tilde{x}_1, \ldots, \tilde{x}_n$ of $\mathbf{x}$ are

Also, given these $n$ samples $\tilde{x}_1, \ldots, \tilde{x}_n$, in a dataset $\mathcal{D}$, find $\theta$. Given the samples $\tilde{x}_1, \ldots, \tilde{x}_n$, in a dataset $\mathcal{D}$, find $\theta$. 

The decision boundary?
\[
\begin{align*}
\text{(Case 1)} & \quad \left( 1 - \frac{p}{n} \right)^{n-1} \left( x - \frac{p}{2} \right)_{\frac{n}{2}} \leq \frac{1}{\sqrt{2 \pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} \\
\text{(Case 2)} & \quad \left( 1 - \frac{p}{n} \right)^{n-1} \left( x - \frac{p}{2} \right)_{\frac{n}{2}} \leq \frac{1}{\sqrt{2 \pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2}
\end{align*}
\]
MAP is similar to ML, except for the extra term $g_n(\theta)$ (prior term).

\[
\frac{p(x|\theta) p(\theta)}{p(x|D)} \frac{1}{(m-1)T} \frac{1}{p(C|D)} = \frac{p(m|\theta, C, T)}{p(C|D)}
\]

Needs to be converted into a posterior $p(\theta|x)$.

In Bayes, $\theta$ is a RV whose distribution of the parameters $\theta$ to be fixed. (Maximize)

In ML (or MAP) estimation, we assume

Bayesian Estimation.
\[
\frac{\partial^2 u}{\partial z^2} + \frac{\partial}{\partial z} \left( \frac{\partial u}{\partial z} \right) = \frac{\partial^2 u}{\partial y^2} + \frac{\partial}{\partial y} \left( \frac{\partial u}{\partial y} \right)
\]

**Note:**

**Gaussian Cases:**

\[
0 \leq \sigma = 0 \leq \sigma (a + b = 1)
\]

\[
\frac{\partial u}{\partial y} = a \frac{\partial u}{\partial y} + b u
\]
For random partitions of the data, repeat the process above. For a CV, leave m/remainder incomplete. Test w/ remainder 10% and tuning the classifier. Cross validation: Take, say, 90% of the data not the real world! Systems model the data. Cons: Biased result because P(O|S) lost of training data. Pros: Use of same data set twice Classifier: Use all data for training. Re-substitution w/ real data
Truth Negative
False Negative
False Positive
True Positive

\[
\begin{array}{c|c|c}
& F & T \\
\hline
F & FN \quad T \quad F \\
\hline
T & TP \quad F \\
\hline
\end{array}
\]

\[\frac{N}{C_1} = 90\% \text{ Correct}\]

Error rates

Confusion Matrix:

\text{Classes:} W_1, W_2, \ldots, W_n

\text{Decision Boundaries:} C_{11}, C_{12}, \ldots, C_{nn}

\text{Decision Rule:} \{x | \text{Class} = C\}

\text{Classification Error:} E(S) = \frac{1}{n} \sum_{i=1}^{n} E_i

\text{Estimation:} E(S) \approx \frac{1}{n} \sum_{i=1}^{n} E_i

\text{True Classes:} W_1, W_2, \ldots, W_n

\text{Estimated Classes:} C_{11}, C_{12}, \ldots, C_{nn}
Recall = \frac{TP}{TP + FN}

Precision = \frac{TP}{TP + FP}

True Positive & Recall Rate
\[ \frac{AVB}{AB} = \frac{TN + FN + FP}{TP} \]

\[ FPR = \frac{TN}{TN + FN} \]

\[ FP + FN \]

\[ TP + FN + FP \]

\[ TP + FN + FP + TN \]

\[ TN + FP \]
Receiver Operating Characteristics

\[ \text{ROC} \]

\[ \frac{\text{TP}}{\text{TP} + \text{FP}} \]

\[ \frac{\text{FP}}{\text{FP} + \text{TN}} \]

\[ \text{FPR} \]

\[ \text{FNR} \]

\[ \text{PPV} \]

\[ \text{NPV} \]
Note: \( \frac{n}{3} \geq 0 \) and \( \frac{n}{\sqrt{3}} \geq 0 \)

\[
\frac{3}{\frac{1}{3}} \left( \frac{3 \frac{n}{1} + 0 \frac{3}{0}}{3 \frac{1}{3}} \right) \geq \frac{n}{\sqrt{3}} + \frac{\sqrt{3}}{\frac{n}{3}} = \frac{3n}{3}.
\]
\[ \frac{c}{\sqrt{|1 - p^2|}} \leq \frac{\sqrt{2}}{\|x \|} \]

In that case, if we make \( \|x\| = 1 = p \):

\[ r^2 = \frac{c}{\sqrt{|1 - p^2|}} \]

\[ \quad \leq \frac{\sqrt{2}}{\|x \|} \]

\( \sqrt{2} \) is diagonal and

\[ x \neq 0 \] or independent (for all \( \neq \)).
The Correspoding EigenVectors of $\Phi^T \Phi$ are the Smallest EigenValues of $\Phi^T \Phi$. In Words: The Minimization is achieved if $\Phi^T \Phi = \Phi = \Phi_{d \times m}$.

By Theorem 4.3.2, we have $\frac{\lambda_i}{\lambda_{\max}} \approx \frac{\| b_i \|^2}{\| b \|^2}$ which we are not using.

In Words: We should replace the $X$,s

Then, $b_{i+1} = b_i - \gamma_i \frac{X}{\lambda_i}$

That is:

The Error at the $i$th Iteration of $X$

The Basis Vector that we want to Minimize

Let's say we want to use only $m > d$ of

Karhunen-Loeve Expansion (K.L. PCA)
Principal Component Analysis

We represent X by projecting it onto the m-first eigenvectors (Φ_m) corresponding to the m-largest eigenvectors (Λ_m).

PCA vs. Class Separation

In this case, the largest (principal) component is certainly not the best one to keep as far as classification goes.