

Análisis Lineal Discriminante

Héctor Selley

Universidad Anáhuac México

13 de junio de 2023

Contenido

- 1 Introducción
- 2 Objetivos
- 3 ¿Qué es el análisis lineal discriminante?
- 4 Función Lineal Discriminante de Fisher
- 5 ¿Cómo funciona el Análisis Lineal Discriminante?

Introducción

- El análisis lineal discriminante (Lineal Discriminant Analysis) fue introducido por el estadístico y biólogo Ronald Aylmer Fisher en el año de 1963.

Introducción

- El análisis lineal discriminante (Lineal Discriminant Analysis) fue introducido por el estadístico y biólogo Ronald Aylmer Fisher en el año de 1963.
- Permite encontrar una combinación lineal de rasgos en común que caracterizan o separan dos o más clases de datos, objetos o eventos.

Introducción

- El análisis lineal discriminante (Lineal Discriminant Analysis) fue introducido por el estadístico y biólogo Ronald Aylmer Fisher en el año de 1963.
- Permite encontrar una combinación lineal de rasgos en común que caracterizan o separan dos o más clases de datos, objetos o eventos.
- LDA está estrechamente relacionado con el análisis de varianza (ANOVA) y el análisis de regresión, ya que ambos buscan expresar una variable independiente como una combinación lineal de características específicas.

Introducción

- El análisis lineal discriminante (Lineal Discriminant Analysis) fue introducido por el estadístico y biólogo Ronald Aylmer Fisher en el año de 1963.
- Permite encontrar una combinación lineal de rasgos en común que caracterizan o separan dos o más clases de datos, objetos o eventos.
- LDA está estrechamente relacionado con el análisis de varianza (ANOVA) y el análisis de regresión, ya que ambos buscan expresar una variable independiente como una combinación lineal de características específicas.
- LDA se encuentra relacionado es con el análisis de componente principal, conocido como PCA, y el análisis factorial.

Introducción

- El análisis lineal discriminante (Lineal Discriminant Analysis) fue introducido por el estadístico y biólogo Ronald Aylmer Fisher en el año de 1963.
- Permite encontrar una combinación lineal de rasgos en común que caracterizan o separan dos o más clases de datos, objetos o eventos.
- LDA está estrechamente relacionado con el análisis de varianza (ANOVA) y el análisis de regresión, ya que ambos buscan expresar una variable independiente como una combinación lineal de características específicas.
- LDA se encuentra relacionado es con el análisis de componente principal, conocido como PCA, y el análisis factorial.
- Es necesario considerar que para la realización del LDA se debe contar con un conjunto de datos que permitan construir un modelo lineal

Introducción

- El análisis lineal discriminante (Lineal Discriminant Analysis) fue introducido por el estadístico y biólogo Ronald Aylmer Fisher en el año de 1963.
- Permite encontrar una combinación lineal de rasgos en común que caracterizan o separan dos o más clases de datos, objetos o eventos.
- LDA está estrechamente relacionado con el análisis de varianza (ANOVA) y el análisis de regresión, ya que ambos buscan expresar una variable independiente como una combinación lineal de características específicas.
- LDA se encuentra relacionado es con el análisis de componente principal, conocido como PCA, y el análisis factorial.
- Es necesario considerar que para la realización del LDA se debe contar con un conjunto de datos que permitan construir un modelo lineal
- La variable dependiente será aquella variable que clasifica a priori una población en subconjuntos de elementos.

Contenido

- 1 Introducción
- 2 Objetivos**
- 3 ¿Qué es el análisis lineal discriminante?
- 4 Función Lineal Discriminante de Fisher
- 5 ¿Cómo funciona el Análisis Lineal Discriminante?

Objetivos

- Generar una combinación lineal de un conjunto de variables explicativas para segmentar una población en subconjuntos determinados por una variable independiente que a priori define la clasificación inicial.

Objetivos

- Generar una combinación lineal de un conjunto de variables explicativas para segmentar una población en subconjuntos determinados por una variable independiente que a priori define la clasificación inicial.
- Identificar las características que diferencian y discriminan a dos o más grupos.

Objetivos

- Generar una combinación lineal de un conjunto de variables explicativas para segmentar una población en subconjuntos determinados por una variable independiente que a priori define la clasificación inicial.
- Identificar las características que diferencian y discriminan a dos o más grupos.
- Generar una función capaz de distinguir con la mayor precisión posible a los miembros de cada grupo.

Contenido

- 1 Introducción
- 2 Objetivos
- 3 ¿Qué es el análisis lineal discriminante?
- 4 Función Lineal Discriminante de Fisher
- 5 ¿Cómo funciona el Análisis Lineal Discriminante?

¿Qué es LDA?

Utilizaremos un ejemplo para explicar en qué consiste el LDA, y su utilidad:

- Supongamos que tenemos un nuevo medicamento experimental para el tratamiento del cáncer.

¿Qué es LDA?

Utilizaremos un ejemplo para explicar en qué consiste el LDA, y su utilidad:

- Supongamos que tenemos un nuevo medicamento experimental para el tratamiento del cáncer.
- Dicho medicamento se ha suministrado a un grupo de pacientes que padecen la enfermedad.

¿Qué es LDA?

Utilizaremos un ejemplo para explicar en qué consiste el LDA, y su utilidad:

- Supongamos que tenemos un nuevo medicamento experimental para el tratamiento del cáncer.
- Dicho medicamento se ha suministrado a un grupo de pacientes que padecen la enfermedad.
- Se ha observado que el medicamento funciona muy bien para algunas personas, mientras que a las demás las ha hecho sentirse peor.

¿Qué es LDA?

Utilizaremos un ejemplo para explicar en qué consiste el LDA, y su utilidad:

- Supongamos que tenemos un nuevo medicamento experimental para el tratamiento del cáncer.
- Dicho medicamento se ha suministrado a un grupo de pacientes que padecen la enfermedad.
- Se ha observado que el medicamento funciona muy bien para algunas personas, mientras que a las demás las ha hecho sentirse peor.
- Esto nos hace plantear la pregunta, ¿cómo se puede decidir a qué personas es conveniente suministrar el medicamento?.

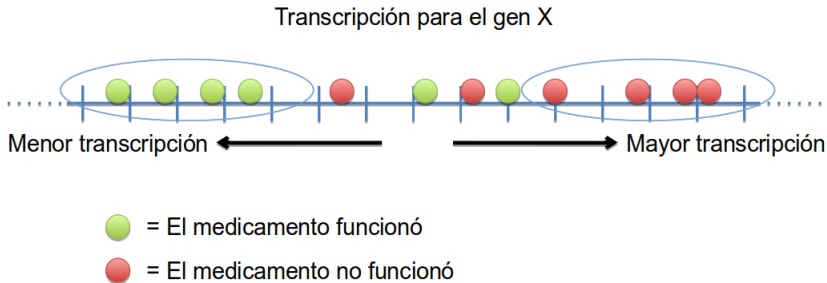
¿Qué es LDA?

Utilizaremos un ejemplo para explicar en qué consiste el LDA, y su utilidad:

- Supongamos que tenemos un nuevo medicamento experimental para el tratamiento del cáncer.
- Dicho medicamento se ha suministrado a un grupo de pacientes que padecen la enfermedad.
- Se ha observado que el medicamento funciona muy bien para algunas personas, mientras que a las demás las ha hecho sentirse peor.
- Esto nos hace plantear la pregunta, ¿cómo se puede decidir a qué personas es conveniente suministrar el medicamento?
- Se utilizará una expresión genética para ayudar a entender el problema y llegar a una posible solución.

¿Qué es LDA?

La Figura muestra una gráfica de la transcripción¹ del gen X, ésta gráfica es una recta numérica que describe la transcripción del gen X, y nos sirve para determinar los pacientes a los cuales se les prescribirá el medicamento y cuáles pacientes no.



¹Transcripción del ADN es el primer proceso de la expresión genética, mediante el cual se transfiere la información contenida en la secuencia del ADN.

¿Qué es LDA?

- En la recta numérica de la Figura anterior, los puntos representan personas.

¿Qué es LDA?

- En la recta numérica de la Figura anterior, los puntos representan personas.
- Mientras más a la izquierda de la recta numérica se encuentre un punto significa que tiene una menor transcripción del gen X.

¿Qué es LDA?

- En la recta numérica de la Figura anterior, los puntos representan personas.
- Mientras más a la izquierda de la recta numérica se encuentre un punto significa que tiene una menor transcripción del gen X.
- Mientras más a la derecha de la recta se encuentre un punto significa que tiene una mayor transcripción del gen X.

¿Qué es LDA?

- En la recta numérica de la Figura anterior, los puntos representan personas.
- Mientras más a la izquierda de la recta numérica se encuentre un punto significa que tiene una menor transcripción del gen X.
- Mientras más a la derecha de la recta se encuentre un punto significa que tiene una mayor transcripción del gen X.
- Los puntos verdes representan un paciente en el cual el medicamento funcionó y los puntos rojos representan un paciente en el cual el medicamento no funcionó.

¿Qué es LDA?

- Se puede observar que para la mayoría de los pacientes con una menor transcripción del gen X el medicamento funcionó, mientras que para la mayoría de los pacientes con una alta transcripción del gen X el medicamento no funcionó.

¿Qué es LDA?

- Se puede observar que para la mayoría de los pacientes con una menor transcripción del gen X el medicamento funcionó, mientras que para la mayoría de los pacientes con una alta transcripción del gen X el medicamento no funcionó.
- En términos generales se puede concluir que el gen X nos indica satisfactoriamente los pacientes que deben tomar el medicamento y aquellos que no, salvo unas excepciones donde este mecanismo no resulta concluyente.

¿Qué es LDA?

- Luego de ver los resultados obtenidos surge la pregunta: ¿cómo se puede mejorar el proceso?.

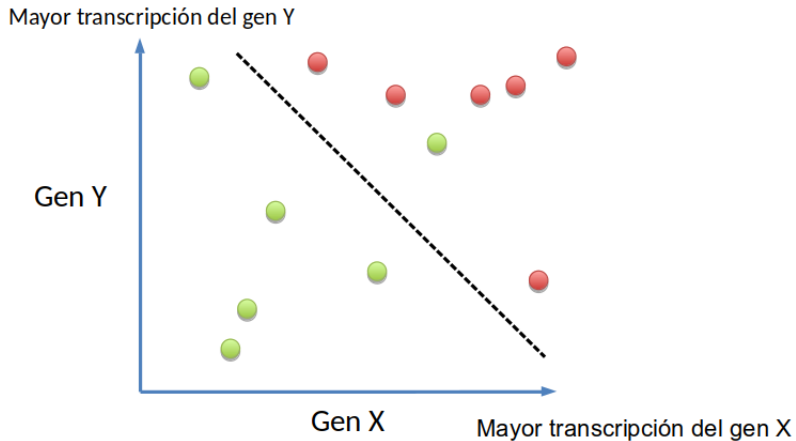
¿Qué es LDA?

- Luego de ver los resultados obtenidos surge la pregunta: ¿cómo se puede mejorar el proceso?.
- Repitamos la idea añadiendo ahora la transcripción del gen Y.

¿Qué es LDA?

- Luego de ver los resultados obtenidos surge la pregunta: ¿cómo se puede mejorar el proceso?.
- Repitamos la idea añadiendo ahora la transcripción del gen Y.
- Esto ocasionará que nuestra gráfica ahora sea en dos dimensiones, donde el eje x corresponde a la transcripción del gen X en los pacientes y el eje y corresponde a la transcripción del gen Y.

¿Qué es LDA?



¿Qué es LDA?

- La línea punteada es una separación entre las dos categorías

¿Qué es LDA?

- La línea punteada es una separación entre las dos categorías
- Los verdes representan los pacientes en los cuales el medicamento funcionó

¿Qué es LDA?

- La línea punteada es una separación entre las dos categorías
- Los verdes representan los pacientes en los cuales el medicamento funcionó
- Los rojos en los cuales el medicamento no funcionó.

¿Qué es LDA?

- La línea punteada es una separación entre las dos categorías
- Los verdes representan los pacientes en los cuales el medicamento funcionó
- Los rojos en los cuales el medicamento no funcionó.
- Es notable que esta separación es mejor que el caso unidimensional, con sólo el gen X

¿Qué es LDA?

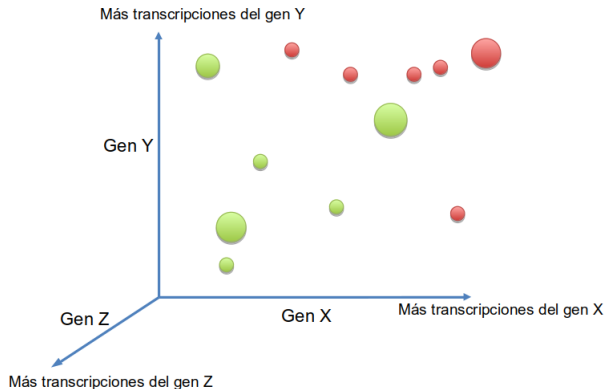
- La línea punteada es una separación entre las dos categorías
- Los verdes representan los pacientes en los cuales el medicamento funcionó
- Los rojos en los cuales el medicamento no funcionó.
- Es notable que esta separación es mejor que el caso unidimensional, con sólo el gen X
- Sin embargo, no todos los pacientes han sido ubicados en la categoría a la que pertenecen.

¿Qué es LDA?

- La línea punteada es una separación entre las dos categorías
- Los verdes representan los pacientes en los cuales el medicamento funcionó
- Los rojos en los cuales el medicamento no funcionó.
- Es notable que esta separación es mejor que el caso unidimensional, con sólo el gen X
- Sin embargo, no todos los pacientes han sido ubicados en la categoría a la que pertenecen.
- Por lo tanto, nuevamente se agrega un gen, en este caso el gen Z, con la finalidad de mejorar la separación de los pacientes.

¿Qué es LDA?

La Figura muestra la gráfica para las transcripciones de los tres genes, observe que ahora es un espacio en tres dimensiones.



¿Qué es LDA?

- En la figura los puntos verdes representan los pacientes en los cuales el medicamento funcionó

¿Qué es LDA?

- En la figura los puntos verdes representan los pacientes en los cuales el medicamento funcionó
- Los puntos rojos aquellos en los que no funcionó.

¿Qué es LDA?

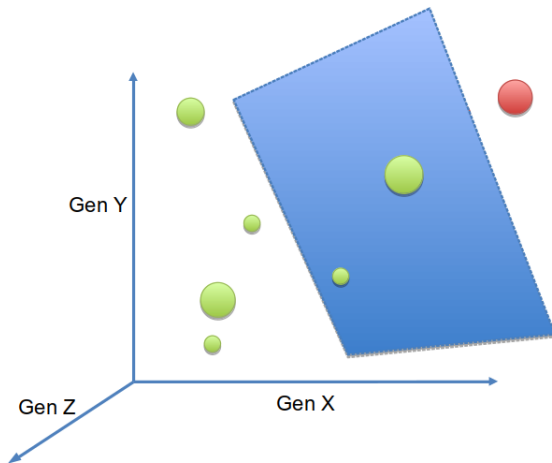
- En la figura los puntos verdes representan los pacientes en los cuales el medicamento funcionó
- Los puntos rojos aquellos en los que no funcionó.
- Se tienen puntos de diferente tamaño lo que significa que tienen una mayor transcripción del gen Z.

¿Qué es LDA?

- En la figura los puntos verdes representan los pacientes en los cuales el medicamento funcionó
- Los puntos rojos aquellos en los que no funcionó.
- Se tienen puntos de diferente tamaño lo que significa que tienen una mayor transcripción del gen Z.
- Para realizar la separación en el espacio tridimensional se requiere de un plano

¿Qué es LDA?

La Figura muestra la separación de los pacientes utilizando un plano.



¿Qué es LDA?

- En esta gráfica se puede observar que el plano logra separar todos los pacientes satisfactoriamente

²La maldición de la dimensión.

¿Qué es LDA?

- En esta gráfica se puede observar que el plano logra separar todos los pacientes satisfactoriamente
- Por lo tanto es el método ideal para realizar esta tarea.

²La maldición de la dimensión.

¿Qué es LDA?

- En esta gráfica se puede observar que el plano logra separar todos los pacientes satisfactoriamente
- Por lo tanto es el método ideal para realizar esta tarea.
- Es natural pensar que mientras más genes se agreguen a los pacientes será mejor para poder realizar una mejor separación de los datos.

²La maldición de la dimensión.

¿Qué es LDA?

- En esta gráfica se puede observar que el plano logra separar todos los pacientes satisfactoriamente
- Por lo tanto es el método ideal para realizar esta tarea.
- Es natural pensar que mientras más genes se agreguen a los pacientes será mejor para poder realizar una mejor separación de los datos.
- Sin embargo, al tratar con más de tres dimensiones, el concepto se complica considerablemente².

²La maldición de la dimensión.

¿Qué es LDA?

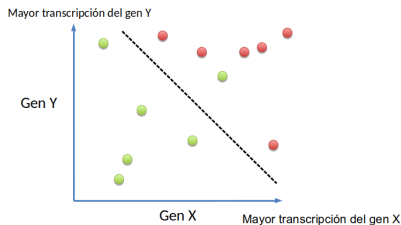
- Lo que se intenta lograr con el Análisis Lineal Discriminante es reducir las dimensiones a solo dos: plano cartesiano xy .

¿Qué es LDA?

- Lo que se intenta lograr con el Análisis Lineal Discriminante es reducir las dimensiones a solo dos: plano cartesiano xy .
- Al realizar el proceso de reducción se buscará maximizar la separación entre los dos grupos, de manera que se puedan tomar mejores decisiones.

¿Qué es LDA?

- Lo que se intenta lograr con el Análisis Lineal Discriminante es reducir las dimensiones a solo dos: plano cartesiano xy.
- Al realizar el proceso de reducción se buscará maximizar la separación entre los dos grupos, de manera que se puedan tomar mejores decisiones.
- Tomemos como punto de partida la Figura y realicemos la reducción de dos a una dimensión.



¿Qué es LDA?

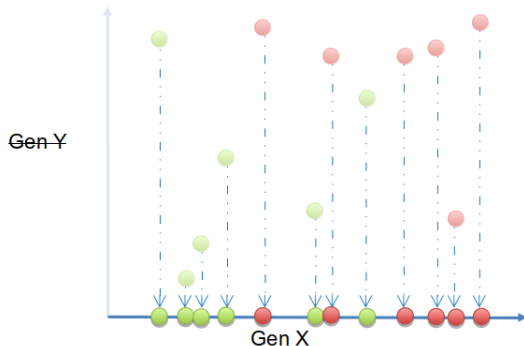
- Una forma muy simple de realizarlo consiste en eliminar por completo el eje del gen Y y proyectar los datos hacia el eje del gen X.

¿Qué es LDA?

- Una forma muy simple de realizarlo consiste en eliminar por completo el eje del gen Y y proyectar los datos hacia el eje del gen X.
- La Figura muestra este procedimiento y el resultado obtenido.

¿Qué es LDA?

- Una forma muy simple de realizarlo consiste en eliminar por completo el eje del gen Y y proyectar los datos hacia el eje del gen X.
- La Figura muestra este procedimiento y el resultado obtenido.



¿Qué es LDA?

- Esta reducción es simple pero se ignora toda la información del gen Y.

¿Qué es LDA?

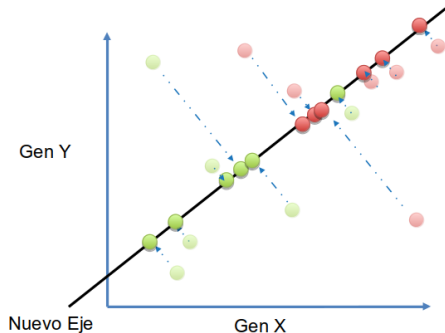
- Esta reducción es simple pero se ignora toda la información del gen Y.
- LDA busca resolver este problema, utilizando la información que proporcionan los dos genes en los dos ejes.

¿Qué es LDA?

- Esta reducción es simple pero se ignora toda la información del gen Y .
- LDA busca resolver este problema, utilizando la información que proporcionan los dos genes en los dos ejes.
- Se define un nuevo eje en el cual se maximice la separación de las dos categorías.

¿Qué es LDA?

- Esta reducción es simple pero se ignora toda la información del gen Y.
- LDA busca resolver este problema, utilizando la información que proporcionan los dos genes en los dos ejes.
- Se define un nuevo eje en el cual se maximice la separación de las dos categorías.



¿Qué es LDA?

- Para LDA se busca maximizar la separación de las dos categorías.

¿Qué es LDA?

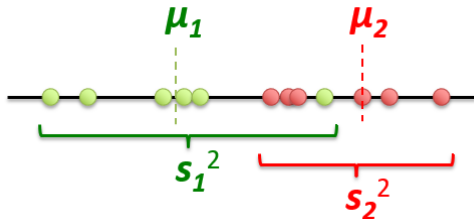
- Para LDA se busca maximizar la separación de las dos categorías.
- Se requiere un nuevo eje mediante el cual se consiga maximizar dicha separación.

¿Qué es LDA?

- Para LDA se busca maximizar la separación de las dos categorías.
- Se requiere un nuevo eje mediante el cual se consiga maximizar dicha separación.
- La forma como mediremos la separación entre los datos es mediante la media aritmética μ , con dispersión (varianza) s^2 .

¿Qué es LDA?

- Para LDA se busca maximizar la separación de las dos categorías.
- Se requiere un nuevo eje mediante el cual se consiga maximizar dicha separación.
- La forma como mediremos la separación entre los datos es mediante la media aritmética μ , con dispersión (varianza) s^2 .



¿Qué es LDA?

- El objetivo de LDA es crear un nuevo eje que permita dividir los datos en dos categorías de acuerdo a dos criterios:

¿Qué es LDA?

- El objetivo de LDA es crear un nuevo eje que permita dividir los datos en dos categorías de acuerdo a dos criterios:
 - ▶ Maximizar la separación de las medias

¿Qué es LDA?

- El objetivo de LDA es crear un nuevo eje que permita dividir los datos en dos categorías de acuerdo a dos criterios:
 - ▶ Maximizar la separación de las medias
 - ▶ Minimizar la dispersión de los datos

¿Qué es LDA?

- El objetivo de LDA es crear un nuevo eje que permita dividir los datos en dos categorías de acuerdo a dos criterios:
 - ▶ Maximizar la separación de las medias
 - ▶ Minimizar la dispersión de los datos

Esto es:

¿Qué es LDA?

- El objetivo de LDA es crear un nuevo eje que permita dividir los datos en dos categorías de acuerdo a dos criterios:
 - ▶ Maximizar la separación de las medias
 - ▶ Minimizar la dispersión de los datos

Esto es:

$$\frac{(\mu_1 - \mu_2)^2}{s_1^2 + s_2^2} = \frac{d^2}{s_1^2 + s_2^2} \quad (1)$$

¿Qué es LDA?

- d se define como la distancia d a la diferencia de las medias

¿Qué es LDA?

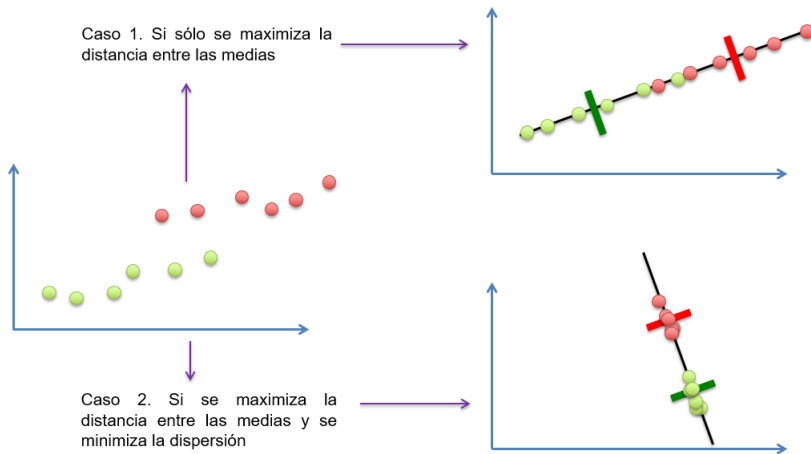
- d se define como la distancia d a la diferencia de las medias
- Se desea que d sea muy grande de manera que la distancia entre las medias sea lo más grande posible.

¿Qué es LDA?

- d se define como la distancia d a la diferencia de las medias
- Se desea que d sea muy grande de manera que la distancia entre las medias sea lo más grande posible.
- Además, se desea que la dispersión de los datos s^2 sea lo más pequeña posible para que las categorías concentren mejor los datos.

¿Qué es LDA?

Resultado de maximizar la distancia de las medias y minimizar la dispersión



Contenido

- 1 Introducción
- 2 Objetivos
- 3 ¿Qué es el análisis lineal discriminante?
- 4 Función Lineal Discriminante de Fisher**
- 5 ¿Cómo funciona el Análisis Lineal Discriminante?

Función Lineal Discriminante de Fisher

- Una matriz X puede hacerse al no asumir una forma paramétrica particular de la distribución de las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$, pero buscando una regla sensible para discriminarlos.

Función Lineal Discriminante de Fisher

- Una matriz X puede hacerse al no asumir una forma paramétrica particular de la distribución de las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$, pero buscando una regla sensible para discriminarlos.
- La sugerencia de Fisher consiste en buscar una función lineal $a'x$

Función Lineal Discriminante de Fisher

- Una matriz X puede hacerse al no asumir una forma paramétrica particular de la distribución de las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$, pero buscando una regla sensible para discriminarlos.
- La sugerencia de Fisher consiste en buscar una función lineal $a'x$
- Debe maximizar el cociente de la suma de los cuadrados de la distancia entre los grupos sobre la suma de los cuadrados de la distancia dentro de los grupos.

Función Lineal Discriminante de Fisher

- Una matriz X puede hacerse al no asumir una forma paramétrica particular de la distribución de las poblaciones $\Pi_1, \Pi_2, \dots, \Pi_g$, pero buscando una regla sensible para discriminarlos.
- La sugerencia de Fisher consiste en buscar una función lineal $a'x$
- Debe maximizar el cociente de la suma de los cuadrados de la distancia entre los grupos sobre la suma de los cuadrados de la distancia dentro de los grupos.

Esto es, sea:

$$y = Xa = \begin{bmatrix} X_1 a \\ \vdots \\ X_g a \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_g \end{bmatrix}$$

Función Lineal Discriminante de Fisher

la combinación lineal de las columnas de X . Entonces y tiene un total de suma de cuadrados

Función Lineal Discriminante de Fisher

la combinación lineal de las columnas de X . Entonces y tiene un total de suma de cuadrados

$$y'Hy = a'X'HXa = a'Ta$$

Función Lineal Discriminante de Fisher

la combinación lineal de las columnas de X . Entonces y tiene un total de suma de cuadrados

$$y'Hy = a'X'HXa = a'Ta$$

los cuales pueden ser particionados como una suma de cuadrados de la distancia dentro de los grupos:

Función Lineal Discriminante de Fisher

la combinación lineal de las columnas de X . Entonces y tiene un total de suma de cuadrados

$$y'Hy = a'X'HXa = a'Ta$$

los cuales pueden ser particionados como una suma de cuadrados de la distancia dentro de los grupos:

$$\sum n_i(\bar{y}_i - \bar{y})^2 = \sum n_i [a'(\bar{x}_i - \bar{x})^2] = a'Ba$$

Función Lineal Discriminante de Fisher

donde \bar{y}_i es la media de los sub-vectores y_i de y , y H es la matriz de tamaño $n_i \times n_i$.
Si a es el vector que maximiza

Función Lineal Discriminante de Fisher

donde \bar{y}_i es la media de los sub-vectores y_i de y , y H es la matriz de tamaño $n_i \times n_i$.
Si a es el vector que maximiza

$$a'Ba/a'Wa$$

se puede denominar a la expresión $a'x$ como *función discriminante lineal de Fisher*.

Función Lineal Discriminante de Fisher

Theorem (Función Lineal Discriminante de Fisher)

El vector a en la función discriminante de Fisher es un eigenvector de $W^{-1} = B$ que corresponde al eigenvalor más grande.

Contenido

- 1 Introducción
- 2 Objetivos
- 3 ¿Qué es el análisis lineal discriminante?
- 4 Función Lineal Discriminante de Fisher
- 5 ¿Cómo funciona el Análisis Lineal Discriminante?

¿Cómo funciona el Análisis Lineal Discriminante?

- La solución propuesta por Fisher es maximizar la función que representa la diferencia entre las medias normalizada por la variabilidad entre clases denominada **dispersión**.

¿Cómo funciona el Análisis Lineal Discriminante?

- La solución propuesta por Fisher es maximizar la función que representa la diferencia entre las medias normalizada por la variabilidad entre clases denominada **dispersión**.
- Para cada clase se define la dispersión como el equivalente a la varianza, esto es la suma del cuadrado de las diferencias entre las muestras proyectadas y la media de su clase correspondiente.

¿Cómo funciona el Análisis Lineal Discriminante?

- La solución propuesta por Fisher es maximizar la función que representa la diferencia entre las medias normalizada por la variabilidad entre clases denominada **dispersión**.
- Para cada clase se define la dispersión como el equivalente a la varianza, esto es la suma del cuadrado de las diferencias entre las muestras proyectadas y la media de su clase correspondiente.

Esto se muestra en la expresión:

$$\tilde{S}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2 \quad (2)$$

¿Cómo funciona el Análisis Lineal Discriminante?

- \tilde{S}_i^2 representa la variabilidad de la clase ω_i luego de proyectarla en su espacio-y.

¿Cómo funciona el Análisis Lineal Discriminante?

- \tilde{S}_i^2 representa la variabilidad de la clase ω_i luego de proyectarla en su espacio-y.
- $\tilde{S}_1^2 + \tilde{S}_2^2$ mide la variabilidad entre las dos clases luego de proyectarlas por lo tanto se llama **dispersión entre clases**.

¿Cómo funciona el Análisis Lineal Discriminante?

- \tilde{S}_i^2 representa la variabilidad de la clase ω_i luego de proyectarla en su espacio-y.
- $\tilde{S}_1^2 + \tilde{S}_2^2$ mide la variabilidad entre las dos clases luego de proyectarlas por lo tanto se llama **dispersión entre clases**.
- El discriminante lineal de Fisher se define como la función lineal $w^T x$ que maximiza la función criterio $J(\omega)$.

¿Cómo funciona el Análisis Lineal Discriminante?

- \tilde{S}_i^2 representa la variabilidad de la clase ω_i luego de proyectarla en su espacio- y .
- $\tilde{S}_1^2 + \tilde{S}_2^2$ mide la variabilidad entre las dos clases luego de proyectarlas por lo tanto se llama **dispersión entre clases**.
- El discriminante lineal de Fisher se define como la función lineal $w^T x$ que maximiza la función criterio $J(\omega)$.
- Esto es la distancia entre las medias proyectadas normalizadas dividido por la dispersión entre clases de los datos, dado por la ecuación:

¿Cómo funciona el Análisis Lineal Discriminante?

- \tilde{S}_i^2 representa la variabilidad de la clase ω_i luego de proyectarla en su espacio-y.
- $\tilde{S}_1^2 + \tilde{S}_2^2$ mide la variabilidad entre las dos clases luego de proyectarlas por lo tanto se llama **dispersión entre clases**.
- El discriminante lineal de Fisher se define como la función lineal $w^T x$ que maximiza la función criterio $J(\omega)$.
- Esto es la distancia entre las medias proyectadas normalizadas dividido por la dispersión entre clases de los datos, dado por la ecuación:

$$J(\omega) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \quad (3)$$

¿Cómo funciona el Análisis Lineal Discriminante?

- Ahora se busca que los datos de la misma clase se proyecten lo más cerca entre sí y además que las medias proyectadas se encuentren lo más lejos posible entre sí.

¿Cómo funciona el Análisis Lineal Discriminante?

- Ahora se busca que los datos de la misma clase se proyecten lo más cerca entre sí y además que las medias proyectadas se encuentren lo más lejos posible entre sí.
- Para encontrar la proyección óptima w^* es necesario expresar $J(w)$ como función de w , por ello se define la dispersión multivariante en forma matricial.

¿Cómo funciona el Análisis Lineal Discriminante?

- Ahora se busca que los datos de la misma clase se proyecten lo más cerca entre sí y además que las medias proyectadas se encuentren lo más lejos posible entre sí.
- Para encontrar la proyección óptima w^* es necesario expresar $J(w)$ como función de w , por ello se define la dispersión multivariante en forma matricial.

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T$$

$$S_w = S_1 + S_2$$

¿Cómo funciona el Análisis Lineal Discriminante?

- Ahora se busca que los datos de la misma clase se proyecten lo más cerca entre sí y además que las medias proyectadas se encuentren lo más lejos posible entre sí.
- Para encontrar la proyección óptima w^* es necesario expresar $J(w)$ como función de w , por ello se define la dispersión multivariante en forma matricial.

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T$$
$$S_w = S_1 + S_2$$

donde S_i representa la matriz de covarianza de la clase w_i y S_w se llama la **matriz de dispersión entre clases** de los datos proyectados.

¿Cómo funciona el Análisis Lineal Discriminante?

La dispersión de la proyección y puede ser representada como una función de la matriz de dispersión, esto es:

¿Cómo funciona el Análisis Lineal Discriminante?

La dispersión de la proyección y puede ser representada como una función de la matriz de dispersión, esto es:

$$\begin{aligned}\tilde{S}_i^2 &= \sum_{y \in \omega} (y - \tilde{\mu}_i)^2 \\ &= \sum_{x \in \omega_i} (w^T x - w^T \mu_i)^2 \\ &= \sum_{x \in \omega_i} w^T (x - \mu_i) (x - \mu_i)^T w \\ &= w^T \left[\sum_{x \in \omega_i} (x - \mu_i) (x - \mu_i)^T \right] w \\ &= w^T S_i w\end{aligned}$$

¿Cómo funciona el Análisis Lineal Discriminante?

y además:

¿Cómo funciona el Análisis Lineal Discriminante?

y además:

$$\begin{aligned}\tilde{S}_1^2 + \tilde{S}_2^2 &= w^T S_1 w + w^T S_2 w \\ &= w^T (S_1 + S_2) w \\ &= w^T S_w w\end{aligned}$$

¿Cómo funciona el Análisis Lineal Discriminante?

Por otro lado, la diferencia entre las medias proyectadas en y se pueden expresar en términos del espacio original x , esto es:

¿Cómo funciona el Análisis Lineal Discriminante?

Por otro lado, la diferencia entre las medias proyectadas en y se pueden expresar en términos del espacio original x , esto es:

$$\begin{aligned}(\tilde{\mu}_1 - \tilde{\mu}_2)^2 &= (w^T \mu_1 - w^T \mu_2)^2 \\&= w^T (\mu_1 - \mu_2) (\mu_1 - \mu_2)^T w \\&= w^T S_B w\end{aligned}$$

donde S_B es la matriz de dispersión entre clases de los datos originales.

¿Cómo funciona el Análisis Lineal Discriminante?

Por otro lado, la diferencia entre las medias proyectadas en y se pueden expresar en términos del espacio original x , esto es:

$$\begin{aligned}(\tilde{\mu}_1 - \tilde{\mu}_2)^2 &= (w^T \mu_1 - w^T \mu_2)^2 \\&= w^T (\mu_1 - \mu_2) (\mu_1 - \mu_2)^T w \\&= w^T S_B w\end{aligned}$$

donde S_B es la matriz de dispersión entre clases de los datos originales. Con estos resultados se puede escribir la ecuación (3) en términos de S_w y S_B , esto es:

¿Cómo funciona el Análisis Lineal Discriminante?

Por otro lado, la diferencia entre las medias proyectadas en y se pueden expresar en términos del espacio original x , esto es:

$$\begin{aligned}(\tilde{\mu}_1 - \tilde{\mu}_2)^2 &= (w^T \mu_1 - w^T \mu_2)^2 \\&= w^T (\mu_1 - \mu_2) (\mu_1 - \mu_2)^T w \\&= w^T S_B w\end{aligned}$$

donde S_B es la matriz de dispersión entre clases de los datos originales. Con estos resultados se puede escribir la ecuación (3) en términos de S_w y S_B , esto es:

$$\begin{aligned}J(w) &= \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \\&= \frac{w^T S_B w}{w^T S_w w}\end{aligned}$$

¿Cómo funciona el Análisis Lineal Discriminante?

$J(w)$ está en función de las medias entre clases normalizado por la matriz de dispersión entre clases. Dado que se busca que la distancia entre las medias de las clases sea máxima, se determinará un máximo de la función $J(w)$.

$$\frac{d}{dw} J(w) = \frac{d}{dw} \left[\frac{w^T S_B w}{w^T S_w w} \right] = 0$$

$$(w^T S_w w) \frac{d}{dw} (w^T S_B w) - (w^T S_B w) \frac{d}{dw} (w^T S_w w) = 0$$

$$(w^T S_w w)(2S_B w) - (w^T S_B w)(2S_w w) = 0$$

$$\left[\frac{w^T S_w w}{w^T S_w w} \right] S_B w - \left[\frac{w^T S_B w}{w^T S_w w} \right] S_w w = 0$$

$$1S_B w - J(w)S_w w = 0$$

$$S_w^{-1} S_B w - J(w)w = 0$$

¿Cómo funciona el Análisis Lineal Discriminante?

Resolviendo el problema para el eigenvalor:

$$S_w^{-1} S_B w = \lambda w \quad \wedge \quad \lambda = J(w)$$

donde λ es un escalar. Entonces esto resulta en que el vector de proyección óptima w^* se puede expresar como:

$$\begin{aligned} w^* &= \text{máx } J(w) \\ &= \text{máx } \left[\frac{w^T S_B w}{w^T S_w w} \right] \\ &= S_w^{-1} (\mu_1 - \mu_2) \end{aligned} \tag{4}$$

¿Cómo funciona el Análisis Lineal Discriminante?

La expresión se denomina como el **discriminante lineal de Fisher**, que determina la dirección de la proyección de los datos en una dimensión. Por lo tanto, la solución será el eigenvector

$$S_w^{-1}S_B$$