

FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE LA REPÚBLICA

INTRODUCCIÓN AL RECONOCIMIENTO DE PATRONES

CURSO 2013

Proyecto final

Autores:

Josué Luiz ÑUÑEZ
Matías TAILANIÁN

Tutores:

Federico LECUMBERRY
Ignacio RAMIREZ



10 de diciembre de 2013

Índice general

1. Introducción	2
2. Base de datos	3
2.1. Características fenotípicas	3
2.2. Características genotípicas - Determinación de polimorfismos (SNPs)	4
2.3. Resumen	4
3. Técnicas utilizadas	6
3.1. Naive Bayes	6
3.2. C 4.5	6
3.3. K-nn	6
3.4. PCA	6
3.5. LDA?	6
3.6. NN?	6
3.7. SVM?	6

Introducción

El objetivo principal del proyecto es la investigación en técnicas que permitan contribuir con la predicción de fertilidad de rodeo lechero y la calidad de la carne integrando técnicas de reconocimiento de patrones sobre datos de alta dimensión.

A lo largo de los años se ha intentado relacionar ciertas características fenotípicas del ganado bovino con algunos indicadores genéticos. Es un problema actualmente abierto de gran interés mundial y Uruguay, siendo un país esencialmente ganadero, no puede estar ajeno. Desde hace varios años se viene desarrollando una línea de investigación relacionada con la genética molecular de la calidad de la carne bovina en el área de Genética de la Facultad de Veterinaria. Se ha realizado un abordaje desde diferentes puntos de vista, como la caracterización de genes conocidos en rodeos vacunos de nuestro país, o la búsqueda y análisis de nuevos genes asociados a la calidad de carne tanto bovina como ovina [1].

Obtener una correlación entre la información genética y la capacidad reproductiva de bovinos, el impacto en la calidad de la leche y de la carne de los animales de nuestros rodeos tendría un alto impacto en la producción. A lo largo de los últimos años en Facultad de Veterinaria se han creado bases de datos con información fenotípica y genotípica relacionada con la calidad cárnica y la fertilidad bovina aplicada a la producción lechera. En esta oportunidad se trabajará con una base de datos confeccionada en nuestro país durante el año 2009, provista por la veterinaria Dra. Ana Meikle.

Base de datos

Para lo que sigue de este trabajo es importante conocer la base de datos con la que se trabajará, y realizar una breve descripción de las características relevantes.

La base de datos consta de información de un seguimiento realizado durante 9 meses sobre 891 vacas de 7 tambos diferentes y consiste en las características detalladas a continuación.

2.1. Características fenotípicas

- Edad.
- Condición corporal (BCS) desde 30 días previos al parto hasta 120 días posparto, con una frecuencia de al menos una vez cada 30 días. Se utilizó la escala de 5 puntos (1=flaca, 5= gorda).
- Cantidad de partos (Cantidad de lactancias). Es un indicador importante que tiene que ver con la historia y desgaste de la vaca.
- Anestro: es la cantidad de días que pasaron desde el parto hasta el reinicio. El reinicio se define como el día en que la progesterona aumenta a un nivel determinado, indicando que la vaca volvió a ciclar.
- Intervalo entre partos. Es la cantidad de días que pasan entre 2 partos consecutivos (válido solamente para vacas multíparas). Para maximizar la producción de leche lo que se busca es que la vaca quede peñada una vez al año, es decir un intervalo entre partos de 365 aproximadamente.
- Secado: la cantidad de días que pasan entre el último día que se le saca leche a la vaca y el parto. Es válido solo para vacas multíparas. Es un indicador de cuanto tiempo se dejó descansar a la vaca antes del parto. Cuanto más tiempo tiene, se prepara el cuerpo y llega en mejor forma.
- Servicios: Cantidad de inceminaciones realizadas para lograr la preñez.
- Concentración de progesterona hasta los 60 días posparto.

- Promedio de cantidad de grasa en la leche durante los primeros 100 días posparto.
- Promedio de cantidad de leche durante los primeros 100 días posparto.

2.2. Características genotípicas - Determinación de polimorfismos (SNPs)

Se extrajo ADN y se obtuvieron muestras de buena calidad. En la figura 2.1 se muestra el procedimiento realizado para caracterizar el gen IGF-I bovino en 3 clases: “AA”, “AB” y “BB”.

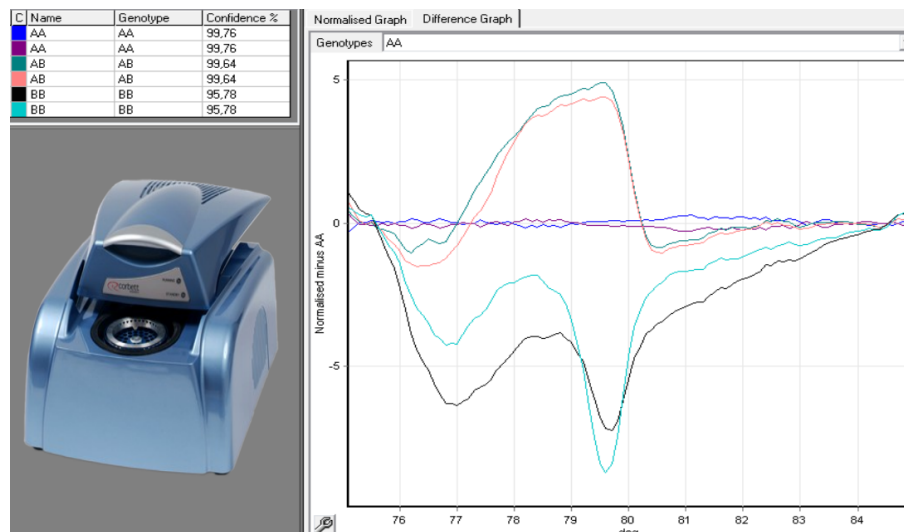


Figura 2.1: Determinación del genotipo

El eje de las ordenadas muestra la fluorescencia normalizada al genotipo “AA”. Se puede observar que el genotipo “AB” se encuentra por encima del genotipo normalizado, mientras que el genotipo “BB” por debajo.

2.3. Resumen

En primer lugar fue necesario entender e interiorizarse con la base de datos, los conceptos y la terminología específica del problema. La lectura e interpretación de los datos no resulta para nada sencilla, por lo que se realizaron dos reuniones con la Dr. Ana Meikle, encargada de la investigación que llevó a cabo la confección de la base asociada a la producción de leche y fertilidad del ganado bovino. En ambas reuniones se buscó depurar la base con el fin de quitar información repetida y no relevante para el estudio. También se estableció un orden de relevancia en las características fenotípicas.

En resumen contamos con una base de datos acotada y “limpia” con varias características fenotípicas que se quieren correlacionar con los genotipos de cada individuo. En particular se abordará el problema como un trabajo de clasificación, tomando los genotipos como clases.

Técnicas utilizadas

En todos los casos se utilizó validación cruzada con 10 subconjuntos.

3.1. Naive Bayes

3.2. C 4.5

3.3. K-nn

3.4. PCA

3.5. LDA?

3.6. NN?

3.7. SVM?

Bibliografía

- [1] Eileen Armstrong Reborati, Detección y análisis de genes asociados a la calidad de la carne en bovinos, *Tesis de doctorado*, Madrid, 2011.
- [2] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11.
- [3] Meyer, K. (2007). WOMBAT – A tool for mixed model analyses in quantitative genetics by REML, J. Zhejiang Uni.